MIT Libraries | DSpace@MIT

MIT Open Access Articles

# *Bayes and Darwin: How replicator populations implement Bayesian computations*

The MIT Faculty has made this article openly available. ***Please share*** how this access benefits you. Your story matters.

**Citation:** Czégel, Dániel, Giaffar, Hamza, Tenenbaum, Joshua B and Szathmáry, Eörs. 2022. "Bayes and Darwin: How replicator populations implement Bayesian computations." BioEssays, 44 (4).

**As Published:** 10.1002/BIES.202100255

**Publisher:** Wiley

**Persistent URL:** https://hdl.handle.net/1721.1/150397

**Version:** Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

**Terms of use:** Creative Commons Attribution NonCommercial License 4.0

Massachusetts Institute of Technology

Received: 28 October 2021 | Revised: 1 February 2022 | Accepted: 3 February 2022

DOI: 10.1002/bies.202100255

BioEssays

PROBLEMS & PARADIGMS

Prospects & Overviews

# Bayes and Darwin: How replicator populations implement Bayesian computations

**Dániel Czégel**[1,2,3,4] | **Hamza Giaffar**[5] | **Joshua B. Tenenbaum**[6,7,8] | **Eörs Szathmáry**[1,2,9]

[1] Institute of Evolution, Centre for Ecological Research, Budapest, Hungary

[2] Parmenides Foundation, Center for the Conceptual Foundations of Science, Pullach, Germany

[3] Doctoral School of Biology, Institute of Biology, Eötvös Loránd University, Budapest, Hungary

[4] Beyond Center for Fundamental Concepts in Science, Arizona State University, Tempe, Arizona, USA

[5] Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, USA

[6] Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

[7] Center for Brains, Minds and Machines, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

[8] Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

[9] Department of Plant Systematics, Ecology and Theoretical Biology, Eötvös Loránd University, Budapest, Hungary

**Correspondence**
Dániel Czégel, Institute of Evolution, ELKH Centre for Ecological Research, 1121 Budapest, Hungary.
Email: danielczegel@gmail.com

**Funding information**
Templeton World Charity Foundation, Grant/Award Number: TWCF0268; Hungarian National Research, Development and Innovation Office - NKFIH, Grant/Award Number: KKP129848

## Abstract

Bayesian learning theory and evolutionary theory both formalize adaptive competition dynamics in possibly high-dimensional, varying, and noisy environments. What do they have in common and how do they differ? In this paper, we discuss structural and dynamical analogies and their limits, both at a computational and an algorithmic-mechanical level. We point out mathematical equivalences between their basic dynamical equations, generalizing the isomorphism between Bayesian update and replicator dynamics. We discuss how these mechanisms provide analogous answers to the challenge of adapting to stochastically changing environments at multiple timescales. We elucidate an algorithmic equivalence between a sampling approximation, particle filters, and the Wright-Fisher model of population genetics. These equivalences suggest that the frequency distribution of types in replicator populations optimally encodes regularities of a stochastic environment to predict future environments, without invoking the known mechanisms of multilevel selection and evolvability. A unified view of the theories of learning and evolution comes in sight.

**KEYWORDS**
adaptation, Bayesian inference, graphical models, particle filters, replicator dynamics

## INTRODUCTION

Learning and evolution both produce adaptive solutions to environmental challenges. The traditional view holds that these processes unfold via distinct mechanisms at different timescales and in different substrates: while evolution happens slowly in populations of replicating individuals through many generations, learning happens within the lifetime of an individual. The two adaptive processes are certainly linked – genetic evolution is the ultimate cause of learning capacities – however learnt information cannot be passed on to other individuals except during social interactions that also evolve. It is legitimate to wonder about the possibility of a deeper link between the two domains,

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.
© 2022 The Authors. *BioEssays* published by Wiley Periodicals LLC

going beyond, for example, the resemblance between operant conditioning and evolution by natural selection. An increasing number of recent studies indeed hint at a deeper link: genetic evolution can store memories of past selective environments and can even generalize, to some extent, to previously unseen ones. There is a deep and fruitful analogy between Bayesian learning and genetic evolution, which seems obvious in hindsight: every hereditary variant is a hypothesis about how to live in the environment: bad genes diminish in frequency just as bad hypotheses diminish in the posterior distribution. This and other insights indicate that a unified view of adaptive processes might be possible.

What is the last word of this sentence? Our mind might be a poor logical inference machine, but it is an excellent probabilistic inference machine.[1,2,3] It evolved to be so: betting hedges over what might happen next has been crucial for survival. Two key ingredients of such a computational process are the representation of hypotheses and the update of their relative plausibility given new evidence. In a computational sense, something strikingly similar happens in evolutionary systems on a different timescale.[4,5] An inherited instruction set builds a phenotype that represents a hypothesis over possible environments and the relative frequencies of phenotypes are updated according to various competitive and cooperative mechanisms. Here we argue that this analogy is both mathematically well-established in terms of the dynamical equivalences of their fundamental models and also conceptually rich in its consequences.

Brains, artificial intelligence algorithms, and biological organisms all face a common challenge: they all need to represent a degree of belief over possible future environments. What is available to make this guess is (a partially observed and noisy version of) the time series of past environments. Both natural environments and complex datasets typically contain many variables (they are high-dimensional) that change over multiple timescales. In such settings, it is not sufficient to simply remember the past; what is needed instead is to extract patterns from the past in order to assign a non-zero probability to environments that have never happened but may yet occur. State of the art generative models, such as generative adversarial networks, do a fairly good job: after being trained with a million images of different bedrooms, they are then able to generate a novel bedroom image that is both coherent and pleasing to the human eye.[6,7] Clearly, the generative model has learned something about the essence of all bedrooms, not just those that it has already encountered. It generalized.

What is the essence of past environments that enables organisms to become adapted to novel environmental conditions?[8] How is this essence stored in heritable form? Theoretical work suggests that natural genetic evolution can permit generalization to a novel environment, provided it is not too novel, that is, it belongs to the same "grammatical class" as the environments to which the ancestral population adapted in the recent past[9]; see also[10] for a general discussion. Our understanding of how biological organisms express inherited information in phenotypes in a possibly environment-dependent way might benefit from the theory of high-dimensional representations as discussed in the fields of artificial intelligence and computational cognitive science.[11,12,13] Here we simply assume that a scalar function over all possible environments is represented in some way, either explicitly or implicitly. In Bayesian learning theory, this is a likelihood function $P(x|h_i)$, which tells us the likelihood of any possible environment (i.e., data) $x$, under hypothesis $h_i$. In the case of organismal phenotypes, it is the fitness $f_i(x)$of type $i$ in environment $x$. With such high-dimensional representations as a starting point, in this paper, we discuss various mathematical and conceptual equivalences of fundamental dynamical models of statistical inference and evolution. How do multiple types, representing different hypotheses, compete? Does an evolutionary system as a whole perform any non-trivial computational function? How efficient is this computation compared to a theoretically optimal one? Statistical learning theory as a mathematical language connects algorithmic (what are local dynamical rules?) and computational (what are global goals that the system as a whole achieves?) level descriptions. By mapping evolutionary dynamics to models of Bayesian statistical learning we also hope to facilitate discussion about various computational functions evolutionary systems might perform in possibly high-dimensional and varying environments. Students of evolution are likely to know about the usefulness of Bayesian statistics through the application of modern phylogenetic methods. What we discuss here is very different – we question to what extent evolution as a process can be regarded as Bayesian, that is probabilistic, computation? Harper[14] raised attention to two striking, formally similar statements. Fisher wrote, *inductive inference is the only process known to us by which essentially new knowledge comes into the world*.[15] As Dawkins asserted fifty years later, *the theory of evolution by cumulative natural selection is the only theory we know of that is in principle capable of explaining the existence of organized complexity*.[16] This article is an argument for the strong epistemological and ontological link between these two observations.

## REPLICATOR DYNAMICS AND BAYESIAN INFERENCE

Logical inference is concerned with inferring the values of deterministic binary variables given the observation of other variables and a model describing the logical relationships between variables. Probabilistic inference generalizes logical inference to update degrees of belief in a model of probabilistic dependencies, not black and white logical values, upon observation of new data. A model of probabilistic dependencies specifies a crucial quantity: the likelihood $P(x|h_i)$ of a specific observation $x$ of observed variables, given a specific setting $h_i$ of hidden variables.

As it turns out, the fundamental dynamics of the process of probabilistic inference and that of replicators with frequency-independent fitness is equivalent once the right identification is established,[17,14,18] see Figure 1. This identification posits that prior beliefs (i.e., prior to observation) over possible values $h_i$ of hidden variables are updated to posterior beliefs (i.e., after observation) just as relative frequencies of types $i$ are updated in discrete-time replicator dynamics. The factor that corresponds to fitness $f_i(x)$ of type $i$ in replicator dynamics is the likelihood $P(x|h_i)$ of an observation $x$, associated with a specific value $h_i$ of the hidden variables. Different settings $h_i$ of hidden variables

15211878, 2022, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bies.202100255 by Massachusetts Institute of Technolo, Wiley Online Library on [04/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

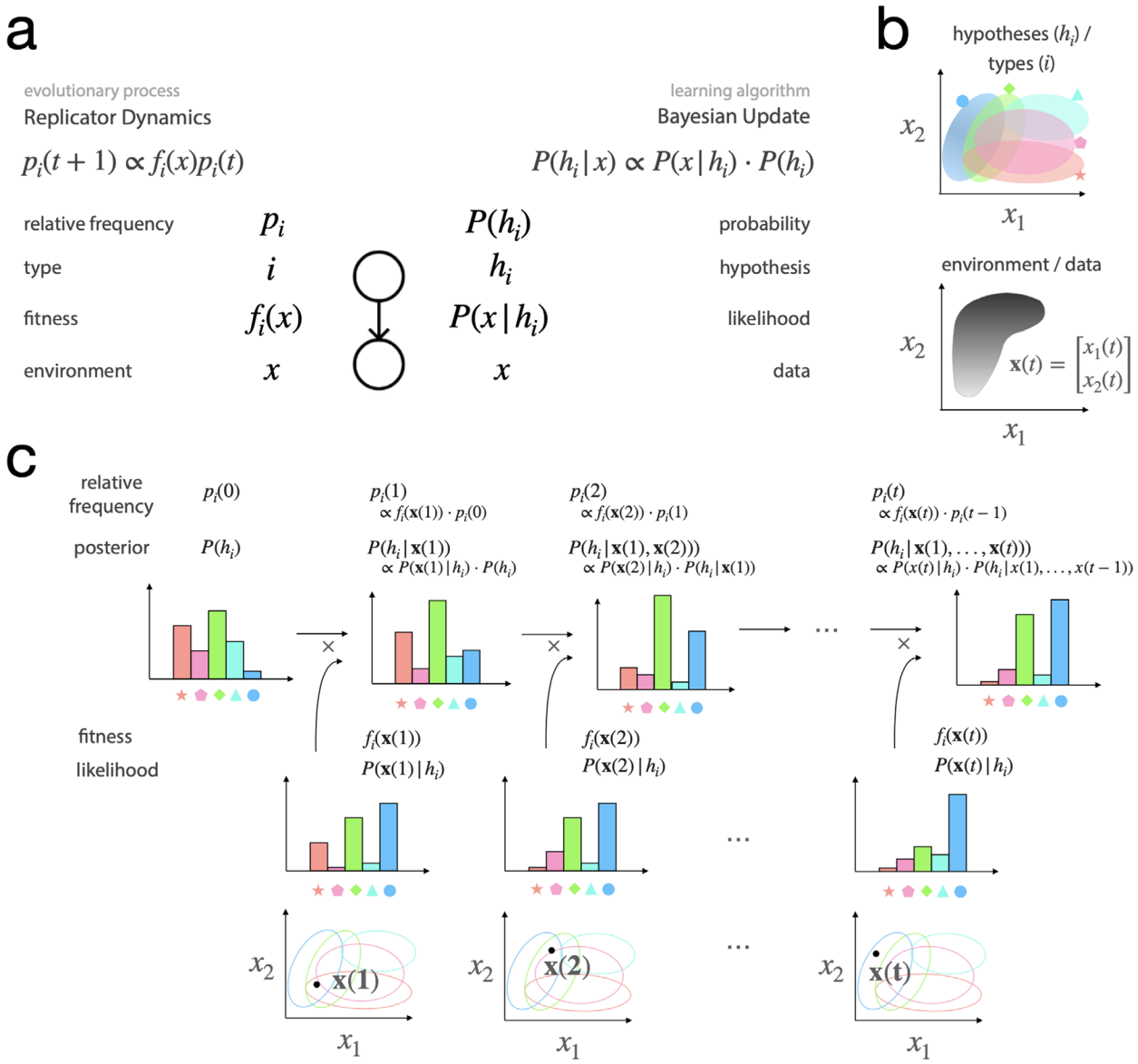


**FIGURE 1** Replicator dynamics and Bayesian update. (A) The graphical model, familiar from Bayesian statistics, corresponding to Bayesian update and to discrete-time replicator dynamics with frequency-independent fitness. Nodes represent either different settings of a random variable with its associated probability distribution or different types with their associated relative frequency distribution. Edges between nodes represent dependencies in a well-defined mathematical sense, see main text for details. (B) Top. Likelihood and fitness functions over data/possible environments, corresponding to the different hypotheses (i.e., settings $h_i$ of the hidden variable $h$) and to the different competing types $i$. Bottom. The parameter space ($x_1$,$x_2$) of data/environment, with its associated probability of occurrence. (C) Dynamics over many sequential steps. Hypotheses that fit the current environment well increase their probability according to Bayes' rule; types that fit the current environment well increase their frequency according to the discrete-time replicator equation

compete, measured in terms of their degree of belief, as types $i$ compete, measured in terms of their relative frequencies. When observing the environment repeatedly, Bayesian update is applied sequentially, according to the rule "today's prior is yesterday's posterior": the posterior at time $t-1$,$P(h_i|x_{t-1},\ldots,x_1)$, is used as the prior at time $t$.

Mathematically, the discrete-time replicator equation with frequency-independent fitness is formulated as

$$p_i(t+1) = \frac{f_i(x_t)p_i(t)}{\sum_i f_i(x_t)p_i(t)} \tag{1}$$

while Bayesian update over hypotheses $h_i$, given data $x_t$, is written as

$$P(h_i \mid x_t, x_{t-1}, \ldots, x_1) = \frac{P(x_t \mid h_i)P(h_i \mid x_{t-1}, \ldots, x_1)}{\sum_i P(x_t \mid h_i)P(h_i \mid x_{t-1}, \ldots, x_1)} \tag{2}$$

In the following, we will not indicate data history $x_t$,$x_{t-1}$,…, $x_1$ explicitly, instead, we focus on the dynamical update at a single timestep $t$. Accordingly, the prior at time $t$, $P(h_i|x_{t-1},\ldots, x_1)$, is simply written as $P(h_i)$. Another simplification in notation is allowed by the fact that the normalization factors, $\sum_i f_i(x)p_i(t)$ and $\sum_i P(x_t|h_i)P(h_i|x_{t-1},\ldots,x_1)$,

15211878, 2022, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bies.202100255 by Massachusetts Institute of Technolo, Wiley Online Library on [04/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

respectively, do not change the ratio of relative frequencies or probabilities. Taken together, replicator dynamics and Bayesian update are written concisely as $p_i(t+1) \propto f_i(x_t)p_i(t)$ and $P(h_i|x_t) \propto P(x|h_i)P(h_i)$.

Two limitations of this analogy, viewed from an evolutionary theory perspective, are the following. First, the replication equation in its general formulation describes frequency-dependent selection: fitness of one type may depend on the frequency of itself as well as on the frequency of any other type.[19] In this sense it is more general than Bayesian update. Second, as it stands, the counterpart of Bayesian learning, is evolution in populations of asexual (clonal), haploid individuals. It is true, however, that many important applications (game theory, ecology) of the replicator equation also ignore this complication: game theory is a model of phenotypic evolution, in which strategies breed.[20] The question how the analogy applies to sexual haploid and diploid populations is open and requires further investigation (cf. will a sexual population evolve to an evolutionarily stable strategy?[21]).

Although in the above discussion we described these dynamics over discrete (geno)types or hypotheses for simplicity, it can be just as well formulated over continuous (pheno)type spaces or continuous parametrizations of hypotheses. In fact, both are used extensively: in Bayesian statistics, this often relies on the existence of a conjugate prior-likelihood pair, which ensures that when the prior is updated to the posterior via the likelihood as described in Equation (2), it stays in the same parametric family (e.g., Gaussian, Dirichlet, etc.). This implies that a series of subsequent Bayesian updates can be viewed as a discrete-time dynamics of a single vector in the parameter space of the prior/posterior (e.g., mean and variance of a Gaussian). Models of (frequency-dependent) phenotypic evolution, such as in game theory and adaptive dynamics, often assume a continuous trait space.[22,23] The population state can be given by a probability measure on a measurable space, and the trait space can have many dimensions.[22] In evolutionary quantitative genetics the dynamics of a continuous phenotypic distribution is being followed under the assumption that phenotypic traits are influenced by many genes, each of them with a small effect on the trait.[24,25] Selection can act on the mean as well as the variance of the population distribution; the genetic variance-covariance matrix plays a crucial role in this approach. Deeper links between evolutionary models in continuous trait spaces and Bayesian models await systematic scrutiny.

## REPLICATOR-MUTATOR DYNAMICS AND FILTERING IN HIDDEN MARKOV MODELS

Mutation between types is introduced to replicator dynamics by allowing probabilistic transitions between types, described by the replicator-mutator equation. Sequential Bayesian update is extended to hidden Markov models (HMMs)[26] the same manner.[27] The mutation kernel between types plays the role of the transition matrix between hidden states (i.e., hypotheses). The replicator-mutator equation takes the form

$$p_i(t+1) = \frac{\sum_j \mu_{i \leftarrow j} f_j(x_t) p_j(t)}{\sum_k f_k(x_t) p_k(t)} \tag{3}$$

Where $\mu_{i \leftarrow j}$ determines the mutation probability (in unit time) from type $j$ to type $i$ and therefore is normalized as $\sum_i \mu_{i \leftarrow j}$, and the factor $\sum_k f_k(x) p_k(t)$ is the average fitness of the population, responsible for keeping the distribution $p_i$ normalized at all times. Equation (3) describes how fitness$f_i(x_t)$ in the current environment, together with the mutation probabilities$\mu_{i \leftarrow j}$, update the relative frequency distribution $p_i$. Similarly to the replicator equation, information about the environment is transferred to the relative frequency distribution over time; here, however, this information is filtered through the mutation probabilities, including the diagonal elements $\mu_{i \leftarrow j}$ that specify the fidelity of replication.

As observed by,[27] this model of replication and mutation is equivalent to updating the probability distribution over latent hypotheses $h_i$ in hidden Markov models (HMMs). HMMs extend Bayesian update by introducing probabilistic transitions between hidden states (i.e., hypotheses) $h_i$, and they infer the joint distribution $\{P(h_i,t=1), P(h_i,t=2), \ldots, P(h_i,t=T)\}$ over the hidden states over time given data history $\{x_1, x_2 \ldots x_T\}$. In order to achieve this, HMMs need two quantities to be pre-determined: hypothesis $h_i$'s likelihood of data, $P(x|h_i)$, as in Bayesian update, and the transition probabilities from $h_j$ to$h_i$, denoted by $P(h_i|h_j)$. This joint probability distribution over hypotheses at all times, $\{P(h_i,t=1), P(h_i,t=2), \ldots, P(h_i,t=T)\}$, can be reduced, however, to distributions of interest; if this distribution of interest is the one corresponding to the last timestep, $P(h_i,t)$, dynamically inferring this last distribution given data history is called the *filtering* problem in HMMs, and it is obtained through the dynamics

$$P(h_i, t+1) = \frac{\sum_j P(h_i|h_j) P(x_t|h_j) P(h_j, t)}{\sum_k P(x_t|h_k) P(h_k, t)} \tag{4}$$

that is indeed equivalent to Equation (3). Importantly, HMMs can also be cast as Bayesian graphical models, as illustrated in Figure 2.

## MULTILEVEL SELECTION AND BAYESIAN INFERENCE IN HIERARCHICAL MODELS

When the replication of different replicator types are partially but not fully synchronized and groups of replicators inherit information regarding the identity of the group, an effective description of the system is provided by multilevel selection theory (MLS).[28] Biological examples, where selection is possibly nonnegligible at multiple levels, include genes within protocells, reproducing organelles in the eukaryotic cell, or individuals in a social insect colony. MLS decomposes the full effect of selection hierarchically to selection acting between and within groups. Partially synchronized replicators form a necessary intermediate step towards new emerging units of evolution, that is, transitions in individuality, that is, in turn, a main mechanism respon-

15211878, 2022, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bies.202100255 by Massachusetts Institute of Technolo, Wiley Online Library on [04/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

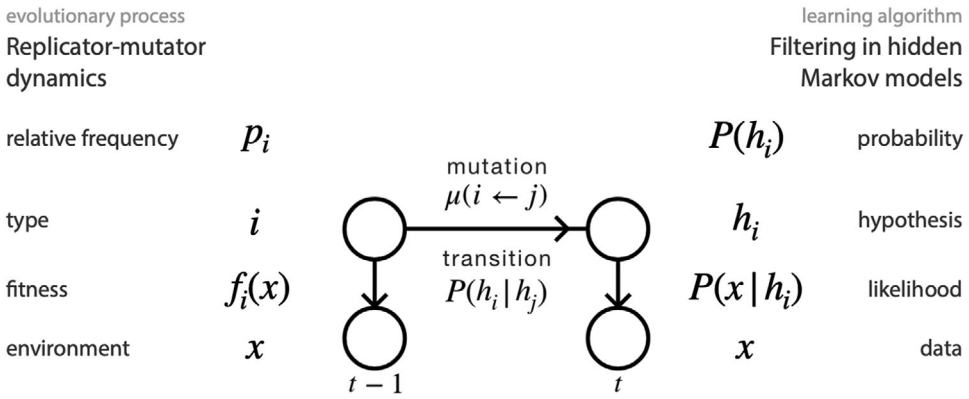


**FIGURE 2** The Bayesian graphical model corresponding to replicator-mutator dynamics and filtering in hidden Markov models (HMMs). Relative frequencies of types are updated by both selection and mutation; the probability distribution over hidden variables (hypotheses) is updated by both their likelihood of fitting current data and transitioning between hidden states. Filtering refers to the fact that the distribution of interest is always the current distribution over hidden variables given data/environment history

sible for increasing complexity in evolutionary systems. Understanding MLS is therefore of crucial importance regarding any, natural or engineered, open-ended Darwinian system.

MLS acts on a hierarchical population of replicators: types of individuals $h_i$ are grouped into types of collectives $z_j$. A complete static description of the population is given by the abundance and composition of collectives (for further details, see[29]). The composition of collective$z_j$, denoted by $c_i^j$, normalized as

$\sum_i c_i^j = 1$, quantifies the relative abundance of individual level replicators within collectives of type $z_j$. The total abundance of collective $z_j$, measured as the total abundance of individual-level replicators within collectives of type$z_j$, is denoted by $p^j$ and normalized as $\sum_j p^j = 1$. These two sets of relative abundances, the compositions of collectives $c_i^j$ and their abundance, $p^j$, describe the composition of the hierarchical (two-level) population completely. From these two quantities, a third one, the total abundance of individual level replicators of type $h_i$ being part of collective $z_j$, denoted by $p_i^j$, can be calculated as$p_i^j = c_i^j p^j$. This quantity connects the two population levels in a sense that the abundance distribution of any level can be computed by summing over the other level: at the level of collectives, the abundance distribution, as we have seen, is given by $p^j = \sum_i p_i^j$; at the level of individuals, the abundance of type $h_i$, $p_i$, is obtained by adding up the abundances of type $h_i$ being in any collective, $p_i = \sum_j p_i^j = \sum_j c_i^j p^j$.

These quantities above describe the composition of the population at one time instance. How does selection change this hierarchical population over time? According to the discrete-time replicator dynamics, abundances change proportionally to their fitness. Here, however, the fitness of an individual-level type $h_i$ might very well depend on the collective $z_j$ it is part of; we denote this fitness by$f_i^j$. We also allow this fitness to depend on the environment $x$. The replicator dynamics, tracking the abundance of individuals of type $h_i$ being part of collectives $z_j$, then reads as

$$p_i^j(t+1) = \sum_j f_i^j p_i^j / \bar{f} \tag{5}$$

where the average fitness, $\bar{f}$, is calculated as $\bar{f} = \sum_{i,j} f_i^j p_i^j$. Tracking the abundance distribution at any level is then possible by summing over the other level, as discussed above. This, however, does not mean that the dynamics at the two levels are decoupled, for example, at the level of individuals, $p_i(t+1) = \sum_j f_i^j p_i^j / \bar{f} \neq (\sum_j f_i^j)(\sum_j p_i^j)/\bar{f}$.

Crucially, fitnesses and abundances are assigned to types that connect the two levels, namely, to individuals of type $h_i$ that are part of collective $z_j$. The replicator equation acts on these quantities, evolving the multilevel population in time. This conceptualization of MLS allows for relating multilevel evolutionary dynamics to hierarchical Bayesian computations over multivariate distributions: these two hierarchical dynamics are structurally equivalent, with the identified quantities listed in Figure 3. Note that this analogy can be extended to arbitrary number of levels, see[29] for details.

An important aspect of multivariate Bayesian models is that it is possible to identify a parametrization independent backbone of the model in terms of conditional independence relations between variables. This allows for distinguishing between model structure (topology) and parameters, a necessary step towards modeling causality.[30] As discussed in,[29] conditional independence relations imposed on the evolutionary implementation of the corresponding Bayesian model have a well-defined intuitive meaning, too: it corresponds to freezing compositions at various levels of the hierarchical population.

Another fundamental feature of hierarchical models is their complexity. Finding the probabilistic model with optimal complexity given data is a non-trivial task that can be approached from many directions. The Bayesian approach is to gauge the model's performance averaged over all possible parameter settings of hidden variables.[31] When comparing two models, possibly having different number of variables and different number of hierarchical levels, one has to favor the one that predicts data better on average. Interestingly, this Bayesian agenda translates to a general and simple evolutionary interpretation: if the average fitness of a collection of replicators is higher when they are

15211878, 2022, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bies.202100255 by Massachusetts Institute of Technolo, Wiley Online Library on [04/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

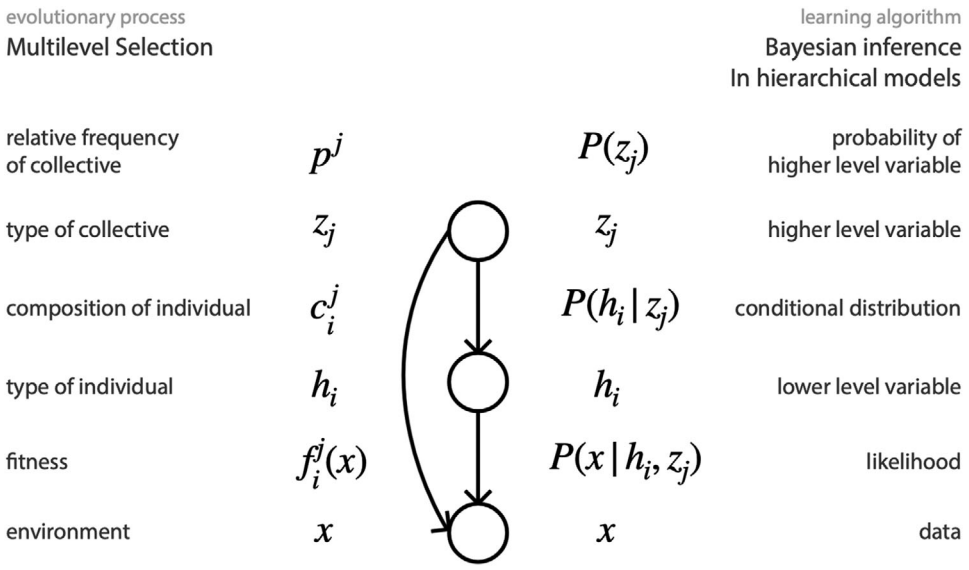


**FIGURE 3** The Bayesian graphical model representing multilevel selection and Bayesian inference in a hierarchical model. Multilevel selection is modeled as discrete-time replicator dynamics over a hierarchical population (types within collectives); in the corresponding hierarchical Bayesian model, two hidden variables fit the data directly and through their dependency structure

grouped together in a collective compared to the case when they are "free", a new level emerges. Although this model of evolutionary transition of individuality[32,33] is rather simplistic in terms of dynamics, it is not in terms of structure, and we believe that this similarity is quite remarkable, having the potential to be refined later on to understand and design complexifying evolutionary systems.

## TIMESCALES OF ADAPTATION

Environmental change might happen at multiple overlapping timescales. A great challenge to any adaptive dynamics is to cover all timescales by various adaptive mechanisms. Evolutionary systems and Bayesian statistical models use (at least) three algorithmically distinct mechanisms.

Developmental/behavioral plasticity and generalization. A fast change from the point of view of a slower adaptive mechanism is perceived as stochastic. In this case, the adaptive mechanism is Bayesian update with or without transition between hidden states (hypotheses), or a replicator dynamic with or without mutation. Hypotheses or types compete to fit all environmental possibilities that might occur at this fast timescale. This includes those that never exactly happened but might happen. Being prepared for such novel environments requires generalization from past experiences. Such adaptedness to all possible environments (weighted by their expected likelihood) is represented by the fitness function $f_i(x)$ or the likelihood function $P(x|h_i)$. Biological organisms might utilize developmental and behavioral plasticity mechanisms, including environment-dependent changes in gene expression to flexibly parametrize $f_i(x)$; Bayesian hypotheses might use various parametrizations of high-dimensional probability distributions, including artificial neural networks, to represent $P(x|h_i)$. Note that fitness and likelihood functions are not adaptive mechanisms; they are the outcome of adaptive mechanisms that operate on a slower timescale.

1. Evolution and tracking data. Environmental change that happens approximately at the timescale of selection and mutation can be tracked by these mechanisms. Replicator-mutator dynamics and (filtering) in hidden Markov models are the simplest models of adaptation to these timescales.
2. Evolvability/facilitated variation and adaptive transition probabilities. Change at timescale 1 can be followed by replicator-mutator dynamics or hidden Markov models, but information contained in the regularities of change is not utilized. Just like prediction at timescale 0 depends on fitting the regularities of this timescale via fitness or likelihood functions, prediction at timescale 1 depends on fitting its regularities via adaptive mutation rates or adaptive transition probabilities between hidden variables.
3. Transitions in individuality and learning the structure of Bayesian models. This is a distinct adaptive mechanism that might bridge multiple timescales. Yet exceeding the limits of selection operating at a single level or of a given structure of a Bayesian model provides opportunity for complexification and for extracting environmental regularities that needs a larger (or different) parameter space to fit. Expanding to new niches and extracting higher-order complex correlations in data are primal examples.

## FROM COMPUTATION TO ALGORITHM: PARTICLE FILTERS

So far, we have discussed computational-level equivalences between learning dynamics in a Bayesian paradigm and infinite-population idealizations of evolutionary dynamics. It is a highly non-trivial

15211878, 2022, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bies.202100255 by Massachusetts Institute of Technolo, Wiley Online Library on [04/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

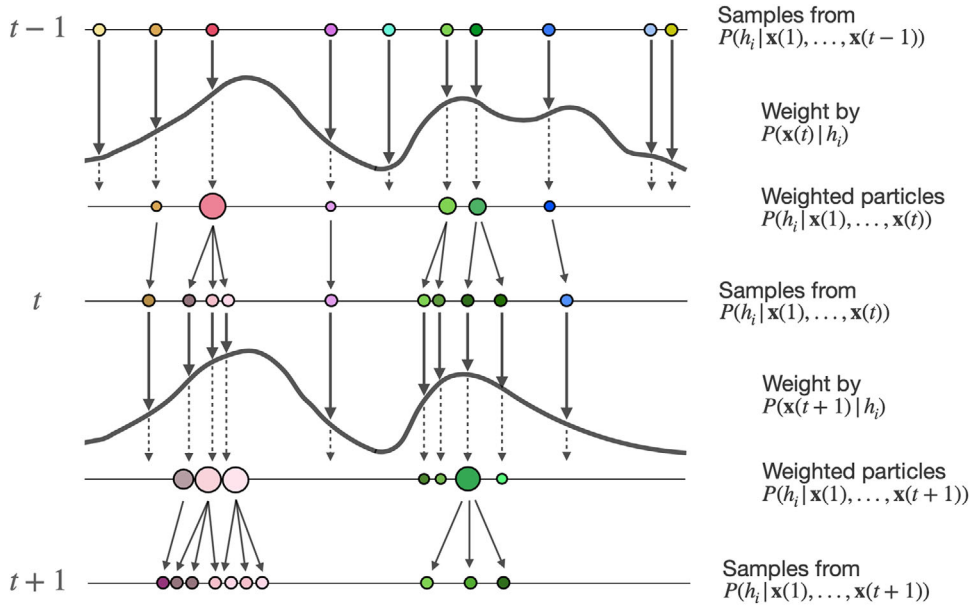


**FIGURE 4** Particle filters and the Wright-Fisher model of population genetics. Computation of weights and resampling collectively correspond to selection and replication. This equivalence implies that the Wright Fisher model implements unbiased sampling of the filtering distribution in hidden Markov models

question whether such equivalences exist between algorithmic-level finite-population approximations as well. The Darwinian algorithm builds on the triplet of replication, variation, and selection over a finite population. It turns out that an equivalent population-based sampling dynamics, particle filters (PFs), are known to sample the filtering distribution in hidden Markov models.[34,35,36] Mathematically, component processes of particle filters map to those of the Wright-Fisher model of population genetics.[18,37] Figure 4 summarizes these component processes and Table 1 presents the equivalences between them.

The critical resampling step of the particle filter was introduced in order to maintain diversity in the population of particles by avoiding particle decay (or blowup). Resampling involves drawing a new set of particles from a weighted distribution such that the normalized weight multiplied by the number of particles is the expected number of times that particle is drawn. The question of when to apply this computationally expensive algorithmic step remains an active research topic in the particle filter field; interestingly, the analogous biological question of when to resample is also non-trivial. Biological evolution rests on multiplication and selectable hereditary variation. Populations cannot persist indefinitely without some form of reproduction; even if organisms were not to age, they would eventually succumb to accidents. This constraint of reproduction does not necessarily apply so simply to components of natural and artificial information processing systems, in which reproduction can be replaced by accumulating scores, as in the assigning of weights to particles or scores to grammatical constructions in Fluid Construction Grammar (FCG),[38] for example. Reproduction becomes necessary when (i) new variants are generated from older ones and (ii) these variants that must be tested against alternatives: template and copy must become distinct. For example, in the case of FCGs, a new grammatical rule must be stored in some new distinct data structure if the ancestor is to be kept as well. In the biological world, clonal plants[39] and animals[40] come closest to the case of accumulating scores. Members of a clonal population, or ramets, do not separate: instead, the colony grows. The size of the colony can be considered more directly analogous to weights or scores. However, the biological constraint remains that colony growth results from reproduction of modules that fail to separate.

Another near ubiquitous mechanism that biological evolution relies on to maintain diversity is genetic recombination. Recombination facilitates adaptation in the sense that biological populations respond faster to directional selection if genes can recombine[41] and permits greater leaps in mutation space. Recombination as an algorithmic element has been introduced to particle filters[42] over a continuous state space, offsprings being set to a weighted average of parent vectors. Genetic recombination, however, suggests that it is testing different combinations of discrete modules, rather than simply mixing parental information, that brings unparalleled advantage over single-parent architectures. A number of evolutionary algorithms have implemented recombination of structured substrates including graphs,[43] matrices[44] and Bayesian graphical models[45,46] in order to find better solutions faster. Perhaps the most sophisticated recombination schemes have been developed for neural networks that rely on modular compressed representations of networks called indirect encodings.[47] In the context of cognition, Monod proposed an evolutionary dynamics of ideas, suggesting that they can also recombine.[48] In the present context we are compelled to wonder about the recombination of hypotheses and how this might proceed all the way from the semantic level down to neurobiological or machine-based implementation?

The mechanical mapping of (replication, variation, selection) to the language of particle filters is crucial in understanding (i) what

15211878, 2022, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bies.202100255 by Massachusetts Institute of Technolo, Wiley Online Library on [04/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

**TABLE 1** Correspondence between particle filters, a sampling-based approximation of posterior distribution in Bayesian statistical learning and the Wright-Fisher model of population genetics

| notations | particle filter | Wright-Fisher model |
|---|---|---|
| $h_i$ | hypothesis/hidden variable $i$ | type $i$ |
| $x_t$ | data at time $t$ | environment at time $t$ |
| $q_t(h_i)$ | probability of hypothesis $i$ at time $t$ | relative frequency of type $i$ at time $t$ |
| $w(h_i,x_t)$ | likelihood of hypothesis $i$ generates data $x_t$ | fitness of type $i$ in environment $x_t$ |
| algorithm | | |
| 1. $\tilde{q}_t(h_i) \leftarrow W(h_i, x_t)q_t(h_i)$ | computing particle weights | selection |
| 2. sample $N$ particles with probability proportional to $\tilde{q}_t(h_i)$ | resampling | replication |
| 3. relocate each particle at $h_i$ to $h_j$ with probability $\mu_{j\leftarrow i}$ | transition between hypotheses | mutation between types |
| 4. $q_{t+1}(h_i) \leftarrow k_i/N$, where $k_i$ is the number of particles at $h_i$ | computing the posterior | next generation relative frequencies |
| limits | | |
| $N \to \infty$ and $\mu \to 0$ | Bayesian update | replicator equation |
| $N \to \infty$ | exact filtering in hidden Markov models | replicator-mutator equation |

statistical computations a population of replicators can perform over their stochastic environment beyond selecting the fittest variant, and (ii) engineering non-parametric statistical learning algorithms based on parallel sampling over potentially non-differentiable, non-parametrized, combinatorial, compositional, and open-ended hypothesis spaces. We note here that another set of constraints might apply: the constraint that the (description of the) hypothesis must be copyable, mutable and evaluable.

## DISCUSSION

Bayesian computations form a unique cornerstone of any adaptive dynamics in uncertain environments. Their power lies in their generality: the algorithmic and implementational details constrained by the given physical/informational substrate is not specified. In this paper we show that it is possible to implement many of the most relevant types of Bayesian computations by simple replicator-based systems.

The possibility of such implementations suggests that replicator systems, although not directly selected for it, perform adaptation in a generalized, system-level, computational sense. It is not only the set of types (or those with highest fitness) that are adapted to the series of past environments; their relative frequencies are maximally informative. Furthermore, they are maximally informative in a predictive sense: given the series of past environments, the system hedges its bets on survival in future environments. Crucially, replicator systems are predictive and perform computations as wholes even without relying on evolvability or multilevel selection. This Bayesian, probabilistic formalization of system-level adaptation of replicator populations can be considered as an extension of previous discussions about the existence of a global cost (energy) function, called the Lyapunov function, over which the dynamics always follows an adaptive path; see Box 1.

Although we think that the general conceptual similarities are possibly far-reaching, there are explicit gaps and mismatches in this analogy. They fall into two categories. Structural mismatches cannot be worked around in the current framework of Bayesian computations, pointing

**Box 1. When do replicator systems perform adaptive computations?**

Finding a Lyapunov function (i.e., a scalar-valued function over the space of parameters on which the learning/evolutionary dynamics always follows a downwards path) is not always easy, finding one out of many possible ones that contributes meaningfully to a learning-theory based interpretation of emergent computations is even harder. In evolutionary theory, this line of thought goes back to Fisher's fundamental theorem of natural selection, stating that a scalar function over (phenotypic) parameters, namely, average fitness, always increases if selection is not frequency-dependent.[62] It is perhaps less well known that a Lyapunov function always exists in frequency-dependent scenarios as well, now interpretable in the language of information theory: it is the KL divergence between the current relative abundance distribution over types and the steady state that always decreases over time.[63,14,64] Following a similar path to find learning theoretically relevant Lyapunov/potential/cost functions in case of more complex dynamics of competing structures that are capable of representing patterns in high-dimensional data is a potentially important step towards a more unified theory of adaptation, incorporating both evolutionary systems composed of competing representations of the environment and learning systems that learn to represent data, in many cases distributed over myriads of elementary computing units interacting locally. A further fundamental conceptual step would be to account for active manipulation of the environment, allowing causal influences between the evolutionary/learning agent and the environment to be mutual. A work connecting niche construction to active inference[50] nicely exemplifies this idea.

15211878, 2022, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bies.202100255 by Massachusetts Institute of Technolo, Wiley Online Library on [04/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

at the need for an extended or modified mathematical framework. One structural mismatch is that fitness functions over all possible environments are non-normalized whereas likelihood functions are normalized (i.e., sum up to 1). Normalization can be regarded as a hard tradeoff between being adapted to different environments; in complex biological reality fitness functions are instead subject to a set of soft physical-chemical-design tradeoffs. Yet as we pointed out in the introduction, the theory of high-dimensional representations (mostly as abstracted from artificial neural networks) offer a promising general direction with possible inbuilt soft or hard tradeoffs regardless of whether the fitness/likelihood function to-be-constructed is normalized or not. Second, likelihood functions associated with Bayesian hypotheses are not explicitly dependent on the prior probabilities of those hypotheses, whereas frequency-dependent fitness plays a key role in modelling the various behavior of replicator systems. The second category, model gaps, fit into the framework of Bayesian statistical learning, but they do not (yet) have exact mathematical analogs. Resource competition can be regarded as such a model gap. Phenomenologically, the Expectation-Maximization algorithm,[49] as used to fit mixture models, builds on the idea of finite shared resources and quasi-equilibrium evolutionary-ecological dynamics. Another conceptual similarity worth mentioning is an added layer to Bayesian computations: the active manipulation of the environment, forming a closed algorithmic loop with building statistical representations of the environment. As pointed out under the specific representational choice of the free energy principle,[4] this closed loop dynamics is analogous to niche construction[50]: the evolved, heritable behavior of causal and directed manipulation of the environment.

When parametrized by neural networks, models of statistical inference share an interesting parallel with development: the separation of timescales. In neural networks, activity patterns distribute information about the stochastic environment at a fast timescale, while synaptic weights extract and store statistical regularities of an ensemble of environments at a much slower timescale. Timescale separation is a simple yet powerful mechanism preventing the interference of the two processes (computation and learning, in the above sense). Fast timescale activity dynamics that compute the consequences of each stochastic environment parallels with developmental plasticity. Indeed, the genotype and current environment jointly determine the phenotype via development. On the other hand, slow timescale plasticity that stores information about regularities of many past environments, maps to evolutionary adaptation encoded in the genotype.

Explicit understanding of developmental encoding of regularities of past environments has been gained using RNA folding and genetic regulatory networks (GRNs) as models of development. RNA folding from its primary sequence can be considered as a simple form of development. The fitness of an RNA structure is given by its structural distance from a target structure that is favored by selection in regime A. Then, for a number of generations, another target B is chosen. During both epochs the RNA population evolves towards the target for a certain number of generations. When we put the RNAs back into regime A, readaptation is faster than that of the naive, primary RNA population. The population shows memory. The population can generalize from the training regimes in that it can adapt fast to a novel target in regime C, provided all target structures belong to the same "grammatical" class.[9] Another example is the memory of genetic regulatory networks (GRNs). The in-silico experiments are similar to the previous RNA case, using alternating selection epochs; target phenotypes are defined by gene activity patterns. Again, it is found that developmental memory evolves.[11] It was also shown that the memory capacity of the GRNs behaves as the memory of analogous neural networks in terms of the number of phenotypes that can be stored, the graceful degradation of memory recall with the progressive removal of regulatory interactions, and the Hebbian-like behavior ('genes that are expressed together have strong mutual, evolved activation') as a population average.[13] A fascinating new research direction is the extent of GRN memory capacity without genetic change by the application of different stimuli.[51]

With overlapping dynamical principles, genetic evolution and Bayesian cognition might be considered as an outer and inner loop of a general adaptive paradigm.[52,53] One interesting question is how they might interact. In particular, whether a Bayesian cognitive mechanism is so powerful that it would hinder genetic evolution—in the extreme, it would render genetic evolution obsolete. In order to be able to do all these calculations better and faster you need classical genetic evolution, including generation and evaluation mechanisms as well as better inherited priors. Once you have them, then genetic evolution will be "allowed" to slow down–at least regarding brain evolution for cognition. As Deacon emphasizes,[54] with a cognitive apparatus hominins were facing relaxed genetic selection in many dimensions, and this may even have facilitated the emergence of complex natural language. Or as Dawkins argued,[55] the main evolutionary players in our species are memes rather than genes.

We envisage three aspects of human inquiry that potentially benefit from this correspondence between evolutionary and probabilistic computations.

First, explaining the adaptive potential of any Darwinian dynamics in Nature: those implemented on a genetic basis, and also those that have not been fully acknowledged, and where the usefulness of replicator-based modeling is questionable at this point, such as memetics,[55,56] Darwinian neurodynamics[57,58,59,60] or quantum Darwinism.[61] Such explanations would mostly be based on either (i) finding global cost functions that evolutionary systems emergently optimize or (ii) relating the computations performed by the system to probabilistic computations that optimally extract information from external data.

Second, many aspects of human and non-human animal cognition is efficiently modeled by the framework of Bayesian inference and generative models. This is in line with selective pressures favoring optimal extraction of information about the environment that in turn enables action selection that leads to maximum survival and reproduction probability. The possible implementations of Bayesian computations on neural substrates, however, is not fully explored. Here we hint at one possible implementation, based on copying and evaluation of any neural information. We do not posit that all Bayesian computations are implemented this way, instead, we point at the possibility that the effective information extraction provided by Bayesian computations

15211878, 2022, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bies.202100255 by Massachusetts Institute of Technolo, Wiley Online Library on [04/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

might be efficiently combined with the open-ended novelty generation evolutionary systems are capable of.

This leads to our third point: the possibility of leveraging the combination of evolutionary and Bayesian computations in designing future AI systems to generate creative yet optimally informed solutions/actions in any probabilistic environment. This direction would, in the very first place, need a more thorough understanding of the relation between evolutionary dynamics and learning theories, possibly unified in the same mathematical language.[10] Indeed, besides highlighting parallels between evolutionary dynamics and learning processes, we believe that one of the most important contributions of this paper is the explicit introduction of Bayesian *graphical* models to evolutionary modeling. Bayesian graphical models provide a consistent modular syntax for combinatorial model building, it unloads the cognitive costs of consistency checks and eases building mental maps of related models. We do not state that all evolutionary processes fit into this framework, neither that the Bayesian graphical framework will be the ultimate language of evolutionary modeling, but we strongly believe that a similar combinatorial syntax could aid focusing on relevant aspects of modeling in evolutionary theory.

## ACKNOWLEDGMENTS

The authors thank István Zachar, Mauro Santos, Gergely Szöllösi, Márton Csillag, Jake Hanson, István Scheuring, Sergey Shuvaev, Sashank Pisupati, Csaba Pléh, András Czégel, and Michael Ollinger for useful discussions and feedback on the manuscript. This work was supported by the National Research, Development and Innovation Office – NKFIH, KKP129848 and by the Templeton Foundation TWCF0268 ("Learning in evolution, evolution in learning"). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## AUTHOR CONTRIBUTIONS

All authors contributed to the concepts, writing and/or revisions of the manuscript.

## ORCID

*Dániel Czégel* https://orcid.org/0000-0002-5722-1598

## REFERENCES

1. Knill, D. C., & Pouget, A. (2004). The bayesian brain: The role of uncertainty in neural coding and computation. *TRENDS in Neurosciences*, *27*(12), 712–719.
2. Körding, K. P., & Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature*, *427*(6971), 244–247.
3. Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science*, *331*(6022), 1279–1285.
4. Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127.
5. Campbell, J. O. (2016). Universal darwinism as a process of bayesian inference. *Frontiers in Systems Neuroscience*, *10*, 49.
6. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. *Advances in Neural Information Processing Systems*, *27*.
7. Goodfellow, I. (2016). Tutorial: Generative adversarial networks. *arXiv preprint arXiv:1701.00160*.
8. Kouvaris, K., Clune, J., Kounios, L., Brede, M., & Watson, R. A. (2017). How evolution learns to generalise: Using the principles of learning theory to understand the evolution of developmental organisation. *PLoS Computational Biology*, *13*(4), e1005358.
9. Parter, M., Kashtan, N., & Alon, U. (2008). Facilitated variation: how evolution learns from past environments to generalize to new environments. *PLoS Computational Biology*, *4*(11), e1000206.
10. Watson, R. A., & Szathmáry, E. (2016). How can evolution learn? *Trends in Ecology & Evolution*, *31*(2), 147–157.
11. Watson, R. A., Mills, R., Buckley, C. L., Kouvaris, K., Jackson, A., Powers, S. T., Cox, C., Tudge, S., Davies, A., Kounios, L., Power, D. (2016). Evolutionary connectionism: Algorithmic principles underlying the evolution of biological organisation in evo-devo, evo-eco and evolutionary transitions. *Evolutionary Biology*, *43*(4), 553–581.
12. Manicka, S., & Levin, M. (2019). The cognitive lens: A primer on conceptual tools for analysing information processing in developmental and regenerative morphogenesis. *Philosophical Transactions of the Royal Society B*, *374*, 20180369.
13. Szilágyi, A., Szabó, P., Santos, M., & Szathmáry, E. (2020). Phenotypes to remember: Evolutionary developmental memory capacity and robustness. *PLoS Computational Biology*, *16*(11):e1008425.
14. Harper, M. (2010). The replicator equation as an inference dynamic. arXiv preprint arXiv:0911.1763.
15. Fisher, R. A. (1937). *The design of experiments*. Oliver & Boyd, Edinburgh.
16. Dawkins, R. (1986). *The blind watchmaker: Why the evidence of evolution reveals a universe without design*. Norton & Company.
17. Cosma Rohilla, S. (2009). Dynamics of bayesian updating with dependent data and misspecified models. *Electronic Journal of Statistics*, *3*, 1039–1074.
18. Suchow, J. W., Bourgin, D. D., & Griffiths, T. L. (2017). Evolution in mind: Evolutionary dynamics, cognitive processes, and bayesian inference. *Trends in Cognitive Sciences*, *21*(7), 522–530.
19. Nowak, M. A. (2006). *Evolutionary dynamics*. Harvard University Press.
20. Smith, J. M. (1982). *Evolution and the Theory of Games*. Cambridge University Press.
21. Smith, J. M. (1981). Will a sexual population evolve to an ess? *The American Naturalist*, *117*(6), 1015–1018.
22. Cressman, R., Hofbauer, J., & Riedel, F. (2006). Stability of the replicator equation for a single species with a multi-dimensional continuous trait space. *Journal of Theoretical Biology*, *239*(2), 273–288.
23. Cressman, R., & Tao, Yi. (2014). The replicator equation and other game dynamics. *Proceedings of the National Academy of Sciences*, *111*(Supplement 3), 10810–10817.
24. Balding, D. J., Bishop, M., & Cannings, C. (2007). *Handbook of statistical genetics*. John Wiley & Sons.
25. Henshaw, J. M., & Zemel, Y. (2017). A unified measure of linear and nonlinear selection on quantitative traits. *Methods in Ecology and Evolution*, *8*(5), 604–614.
26. Ghahramani, Z. (2001). An introduction to hidden markov models and bayesian networks. In *Hidden Markov models: applications in computer vision* (*pp* 9–41). World Scientific.
27. Akyıldız, O. D. (2017). A probabilistic interpretation of replicator-mutator dynamics. arXiv preprint arXiv:1712.07879.
28. Okasha, S. (2001). Why won't the group selection controversy go away? *The British Journal for the Philosophy of Science*, *52*(1), 25–50.
29. Czégel, D., Zachar, I., & Szathmáry, E. (2019). Multilevel selection as bayesian inference, major transitions in individuality as structure learning. *Royal Society Open Science*, *6*(8), 190202.

15211878, 2022, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bies.202100255 by Massachusetts Institute of Technolo, Wiley Online Library on [04/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

30. Pearl, J. (2009). *Causality*. 2, Cambridge University Press.
31. MacKay, D. J. C. (2003). Information theory, inference, and learning algorithms.
32. Smith, J. M., & Szathmáry, E. (1995). *The major transitions in evolution*. Freeman, Oxford.
33. West, S. A., Fisher, R. M., Gardner, A., & Toby Kiers, E. (2015). Major evolutionary transitions in individuality. *Proceedings of the National Academy of Sciences*, *112*(33), 10112–10119.
34. Gordon, N. J., Salmond, D. J., & Smith, A. F. M. (1993). Novel approach to nonlinear/non-gaussian bayesian state estimation. In *IEE Proceedings F-radar and signal processing*, volume *140*, pages 107–113. IET.
35. Doucet, A., Godsill, S., & Andrieu, C. (2000). On sequential monte carlo sampling methods for bayesian filtering. *Statistics and computing*, *10*(3), 197–208.
36. Kwok, N. M., Fang, G., & Zhou, W. (2005). Evolutionary particle filter: re-sampling from the genetic algorithm perspective. In *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2935–2940. IEEE.
37. Griffiths, T. L. (2020). Understanding human intelligence through human limitations. *Trends in Cognitive Sciences*.
38. Steels, L., & Szathmáry, E. (2016). Fluid construction grammar as a biological system. *Linguistics Vanguard*, *2*(1).
39. Pan, J. J., & Price, J. S. (2002) Fitness and evolution in clonal plants: the impact of clonal growth. In *Ecology and evolutionary biology of clonal plants* (pp 361–378). Springer.
40. Simpson, C., Herrera-Cubilla, A., & Jackson, J. B. C. (2020). How colonial animals evolve. *Science Advances*, *6*(2), eaaw9530.
41. Smith, J. M. (1989). *Evolutionary genetics*. Oxford University Press.
42. Uosaki, K., Kimura, Y. & Hatanaka, T. (2003). Nonlinear state estimation by evolution strategies based particle filters. In The 2003 Congress on Evolutionary Computation CEC'03., volume 3, pages 2102–2109. IEEE.
43. Olariu, S., & Zomaya, A. Y. (2005). *Handbook of bioinspired algorithms and applications*. CRC Press.
44. Wallet, B. C., Marchette, D. J., & Solka, J. L. (1996). Matrix representation for genetic algorithms. In *Automatic Object Recognition VI* (volume 2756, *pp* 206–214). International Society for Optics and Photonics.
45. Myers, J. W., Laskey, K. B., & DeJong, K. A. (1999). Learning bayesian networks from incomplete data using evolutionary algorithms. In *Proceedings of the 1st Annual Conference on Genetic and Evolutionary Computation-Volume 1*, pages 458–465.
46. Muruzábal, J., & Cotta, C. (2007). A study on the evolution of bayesian network graph structures. In *Advances in Probabilistic Graphical Models* (*pp* 193–213). Springer.
47. Stanley, K. O., D'Ambrosio, D. B., & Gauci, J. (2009). A hypercube-based encoding for evolving large-scale neural networks. *Artificial Life*, *15*(2), 185–212, 04.
48. Monod, J. (1971). Chance and necessity: an essay on the natural philosophy of modern biology.
49. Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.
50. Constant, A., Ramstead, M. J. D., Veissiere, S. P. L., Campbell, J. O., & Friston, K. J. (2018). A variational approach to niche construction. *Journal of The Royal Society Interface*, *15*(141), 20170685.
51. Biswas, S., Manicka, S., Hoel, E., & Levin, M. (2021). Gene regulatory networks exhibit several kinds of memory: Quantification of memory in biological and random transcriptional networks. *Iscience*, *24*(3), 102131.
52. Geisler, W. S., & Diehl, R. L. (2002). Bayesian natural selection and the evolution of perceptual systems. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *357*(1420), 419–448.
53. Lo, A. W., & Zhang, R. (2021). The evolutionary origin of bayesian heuristics and finite memory. *Iscience*, *24*(8), 102853.
54. Deacon, T. W. (2010). A role for relaxed selection in the evolution of the language capacity. *Proceedings of the National Academy of Sciences*, *107*(Supplement 2), 9000–9006.
55. Dawkins, R. (1976). *The selfish gene*. Oxford University Press.
56. Aunger, R. (2002). *The electric meme: A new theory of how we think*. Simon and Schuster.
57. Fernando, C. T., Szathmáry, E., & Husbands, P. (2012). Selectionist and evolutionary approaches to brain function: A critical appraisal. *Frontiers in Computational Neuroscience*, *6*(24).
58. Szilágyi, A., Zachar, I., Fedor, A., de Vladar, H. P., & Szathmáry, E. (2016). Breeding novel solutions in the brain: A model of darwinian neurodynamics. *F1000Research*, *5*(2416).
59. Fedor, A., Zachar, I., Szilágyi, A., Ollinger, M., de Vladar, H. P., & Szathmáry, E. (2017). Cognitive architecture with evolutionary dynamics solves insight problem. *Frontiers in Psychology*, *8*.
60. Czégel, D., Giaffar, H., Csillag, M., Futó, B., & Szathmáry, E. (2021). Novelty and imitation within the brain: A darwinian neurodynamic approach to combinatorial problems. *Scientific Reports*, *11*(1), 1–14.
61. Hubert Zurek, W. (2009). Quantum darwinism. *Nature Physics*, *5*(3), 181–188.
62. Fisher, R. A. (1930). *The genetical theory of natural selection*. The Clarendon Press.
63. Akin, E. (1979). *The geometry of population genetics*. Springer-Verlag, Berlin.
64. Baez, J. C., & Pollard, B. S. (2016). Relative entropy in biological systems. *Entropy*, *18*(46).

**How to cite this article:** Czégel, D., Giaffar, H., Tenenbaum, J. B., & Szathmáry, E. (2022). Bayes and Darwin: How replicator populations implement Bayesian computations. *BioEssays*, *44*, e2100255. https://doi.org/10.1002/bies.202100255

15211878, 2022, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bies.202100255 by Massachusetts Institute of Technolo, Wiley Online Library on [04/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License