

**Acceleration of evolutionary processes by learning and extended Fisher's fundamental theorem**So Nakashima<sup>1,\*</sup> and Tetsuya J. Kobayashi<sup>1,2,3,†</sup><sup>1</sup>*Faculty of Mathematical Informatics, Graduate School of Information Science and Technology,  
University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8654, Japan*<sup>2</sup>*Institution of Industrial Science, University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan*<sup>3</sup>*Universal Biology Institute, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8654, Japan*

(Received 8 June 2021; accepted 4 January 2022; published 31 January 2022)

Natural selection is a general and powerful concept to explain evolutionary processes in biological organisms and to design engineering systems such as genetic algorithms. In conventional biology, it is assumed that changes in an organism occur solely from random mutations, which are independent of its ancestors' experiences. However, there is accumulating evidence that organisms transmit their information to the next generation via epigenetic states or other methods of information transfer. This information may enable descendants to learn adaptive traits without relying on selection and may accelerate evolution when combined with natural selection. The combination of natural selection and individual learning is also becoming important for engineering applications to improve genetic algorithms or reinforcement learning. While acceleration of the evolutionary process by learning has previously been suggested, no theoretical foundation is available to support it. To accelerate evolutionary processes by learning, individuals should be able to learn behaviors to optimize fitness, which is a trait of the population rather than the individual. It has not yet been clarified whether and how individual learning can improve fitness and thereby accelerate evolution. We also lack a methodology to quantify acceleration and thus enable us to understand, verify, and predict the impacts of learning. In this work, we show that agents can accelerate the evolutionary process by ancestral learning, which employs the information transmitted only from their ancestors (ancestral information). We then clarify that acceleration occurs because ancestral information enables agents to estimate the gradient of fitness. Finally, to quantify acceleration, we extend Fisher's fundamental theorem for natural selection to ancestral learning. The extended Fisher's fundamental theorem relates acceleration to the variation of individual fitness, and thus enables a quantitative understanding of when, how, and why learning is beneficial.

DOI: [10.1103/PhysRevResearch.4.013069](https://doi.org/10.1103/PhysRevResearch.4.013069)**I. INTRODUCTION**

A fundamental question in evolutionary biology is how organisms acquire sophisticated traits, functions, and strategies to survive in harsh and ever-changing environments. Attempts to answer this question have led to the development of the theory of natural selection [1]. The evolutionary process by natural selection is general and powerful enough not only to explain various biological phenomena but also to be applied to mathematical optimization in applied mathematics. Genetic and evolutionary algorithms [2] are used for mathematical optimization by simulating the “evolution” of candidate solutions. Natural selection is versatile and powerful as it does not require active information processing of individual organisms or agents. Due to the growth competition among a population,

the nondirectional random mutations in individual organisms show a trend toward higher fitness at the population level. This can explain biological evolution without the need for advanced or purposeful design. Furthermore, it can work as a basis for optimization algorithms in engineering, such as genetic and evolutionary algorithms, that require small efforts of fine tuning for specific problems. In this sense, conventional natural selection is both a passive and collective process.

However, several studies in biology and engineering have recently considered the natural selection of intelligent agents [Fig. 1(a)], which can learn from experience and actively change their traits accordingly. In biological systems, the gain of fitness by sensing environmental information has previously been investigated [3–6]. Some studies [7,8] indicated that there was a possibility that learning may accelerate the evolutionary process. While this idea seems to violate the conventional presumption that changes in an organism result solely from random mutations that are independent of its ancestors' experiences, there is accumulating evidence that organisms can transmit information to the next generation using genes as well as epigenetic states or by other means such as culture [9]. The transmitted information may thus be employed by descendants to learn and accelerate evolution. This point of view may be important for understanding the

\*so\_nakashima@mist.i.u-tokyo.ac.jp

†tetsuya@mail.crmind.net

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

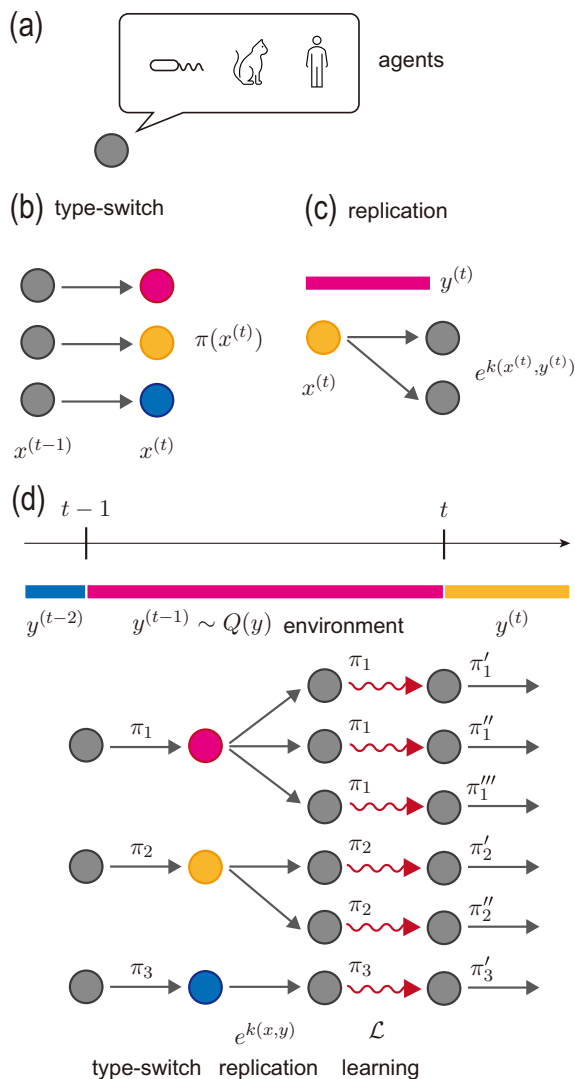


FIG. 1. Schematic representation of the setup used for learning in evolutionary processes. (a) Agents that can replicate and learn. Examples include microbes, animals, and humans. (b)–(d) Schematic illustrations of the model. (b) Type switching  $\pi$  of the agents. An agent at time  $t - 1$  first determines its type based on the strategy  $\pi$ . The colors of the agents other than gray represent the expressed types. The color of the environment represents the state. (c) Replication of the agents. An agent produces  $e^{k(x, y^{(t-1)})}$  daughters, depending on the type  $x$  and the environmental state  $y^{(t-1)}$ . The environmental state  $y^{(t-1)}$  independently follows  $Q(y)$ . (d) Learning by the agents. After replication, the daughters inherit strategies updated by a given learning rule  $\mathcal{L}$ .

fitness value of the phenotypic variability of organisms and their correlations over generations, which are measured by recently developed experimental techniques [10–16].

A pioneering study by Xue and Leibler [7] considered a growing population of agents, each of which follows a learning rule to choose the same type as that of its parent more frequently than the parent (referred to as Xue’s rule). They showed that this simple learning rule could acquire the optimal type-switching strategy in a changing environment.

In engineering, it has been shown that genetic and evolutionary algorithms can be improved by introducing learning by individual agents. A memetic algorithm [17] and information geometric optimization [18] are examples of extended optimization algorithms that employ an active update of candidate solutions by, for example, gradient descent.

Consequently, there has been a surge of interest, both in biology and engineering, regarding the interplay between natural selection and individual learning. Although this idea has been suggested and discussed in both fields, the theoretical basis has not yet been established.

### Learning in evolutionary processes

As the interplay between natural selection and individual learning is tangled, we first describe the situation and the definition of learning before clarifying the problems.

We consider a population of agents that asexually replicate. Each agent has a type and stochastically selects one type in one generation. The type can affect the number of offspring that the agent can generate. In biological systems, the type can be interpreted as a phenotypic trait of an organism. For engineering systems, e.g., genetic algorithms, the type represents a candidate solution for the optimization problem. The type cannot be directly inherited between generations. Each agent also has a type-switching strategy that determines the probability of choosing each type. Stochastic type selection can be beneficial when the state of the environment changes over time [19,20]. We assume that the strategy is heritable and subject to selection. From a biological viewpoint, this strategy can be regarded as a genetic or epigenetic trait, and the types (phenotypes) of agents can be correlated among generations via the inheritance of the strategy. From an engineering viewpoint, this strategy is related to hyperparameters that determine the behavior of an agent.

Since the strategy is heritable, better strategies can be selected via natural selection if there is a diversity of strategies in a population. In a conventional evolutionary process, the diversity of strategies is generated by random (mutational or epigenetic) changes that occur when the strategy of an individual agent is inherited from one generation to the next. We define learning of individual agents as the situation in which the inherited strategy is biased based on the past information of ancestors or the population. Specifically, we consider learning rules that bias the offspring strategy to gain greater fitness. In general, conventional random changes can also be regarded as kinds of learning rules in which no average gain of fitness is expected. Therefore, we call them passive or zero-order learning rules. The main focus of this work is on the learning rules that can bias the strategy so as to achieve an average gain of fitness. Agents should update the strategy in the direction called a *gradient* of fitness towards which the fitness increases (to be defined rigorously in Sec. V). We call them active or first order because the gradient is closely related to the first derivative of the fitness with respect to the strategy [21].

Under the above settings, there are three main problems regarding the interplay between natural selection and learning. The first is whether or when learning can accelerate the evolutionary process of agents to acquire the optimal strategy. Since a learning rule must be simple enough to be implemented in

biological systems, we should investigate whether the evolutionary process is accelerated, even with simple learning rules. For engineering systems, such simplicity is desirable to build a scalable learning algorithm. In a previous study by Xue and Leibler [7], Xue's rule was shown to achieve the optimal strategy under a constant environment via an evolutionary process. However, zero-order random changes can also achieve the optimal strategy; therefore, active learning may not always be beneficial or efficient when compared with passive learning.

The second problem is whether an agent can estimate the gradient from the accessible information. Specifically, we do not know what kind of information is sufficient to estimate the gradient. Although Xue's learning rule can find the optimal strategy by using only the information of the parent's type, the relationship between Xue's rule and the fitness gradient is unclear. The information of the parent's type might be insufficient to estimate the gradient. Since fitness is a trait of the population, communication among agents of the same generation might be required to estimate the fitness gradient. Sufficient conditions are also important for engineering systems to find new variants of genetic and evolutionary algorithms.

The last is how to quantify and predict the acceleration of natural selection by learning. For conventional evolutionary processes with natural selection, we use Fisher's fundamental theorem (FF-thm) and its variants [22]. The theorem states that the increase in the mean fitness of a population is proportional to the variance of fitness in the population. From this relationship, we can predict the progress and speed of evolution in a population. As the evolutionary process becomes more complicated by taking the learning of individual agents into account, a simple relationship similar to FF-thm would help to facilitate our understanding of the impact and efficiency of learning. Furthermore, such a relationship can be applied to analyze the performance of engineering systems.

In this paper, we address these three problems. First, we propose *ancestral learning*, which utilizes only the information transmitted from the ancestors. Ancestral learning is simple and therefore biologically reasonable, which also generalizes Xue's learning rule. The candidates of the information carrier for ancestral learning include the abundance of proteins and mRNAs [10], epigenetic scars of social defeat stress [11], gut microbiota [12], and the intergenerational effects of space flight on epigenetic states [13]. We validate that learning accelerates the evolutionary process, by numerically showing that the fitness with ancestral learning increases faster than with the zero-order mutational rule. Second, we prove that the ancestral information is sufficient to estimate the fitness gradient. We also clarify that ancestral learning is the first-order learning rule by proving that ancestral information is sufficient to estimate the fitness gradient. Third, we derive an extended FF-thm for ancestral learning, which relates the variation in fitness among ancestors to the fitness gain by ancestral learning. Using this theorem, we can predict the acceleration of evolutionary processes by ancestral learning, which depends on the properties of the environment. The theorem enables us to quantitatively understand when and why ancestral learning becomes beneficial.

## II. MATHEMATICAL MODELS AND SETUP

We consider the population dynamics of asexual agents with a discrete generation time  $t \in \{0, 1, 2, \dots\}$ . Let  $x^{(t)} \in \mathcal{X}$  and  $y^{(t)} \in \mathcal{Y}$  be the type of an agent and the state of the environment at time  $t$ . The type models the phenotypic traits of organisms in biological systems. Each agent has its own stochastic type-switching strategy  $\pi_F \in \mathbb{R}^{\mathcal{X}}$  where  $\pi_F(x)$  is the probability to switch into type  $x$  and  $\pi_F$  satisfies that  $\sum_{x \in \mathcal{X}} \pi_F(x) = 1$  and  $\pi_F(x) \geq 0$  for all  $x \in \mathcal{X}$  [Fig. 1(b)]. We call  $\pi_F$  a strategy and  $y^{(t)} \in \mathcal{Y}$  the environmental state at time  $t$ . The environmental state  $y^{(t)}$  at each time  $t$  independently follows the same distribution  $Q(y)$  on  $\mathcal{Y}$ . An agent with type  $x$  under the environmental state  $y$  duplicates asexually and produces  $e^{k(x,y)}$  daughters on average [Fig. 1(c)]. The term  $e^{k(x,y)}$  is called an *individual fitness* [23] of the agent. We define paths (histories) of the types along a lineage and environmental states from time zero to time  $t - 1$  as  $\mathbb{X}^{(t)}$  and  $\mathbb{Y}^{(t)}$ , respectively.

To define the "fitness" of a strategy, we first consider a case in which the agents cannot learn the strategy and the strategy  $\pi_F$  is fixed in a population and over generations. The number  $N^{(t)}$  of agents in the population at time  $t$  under the path  $\mathbb{Y}^{(t)}$  of the environmental states becomes as follows:

$$N_{\pi_F}^{(t)}[\mathbb{Y}^{(t)}] = \left[ \sum_{x \in \mathcal{X}} e^{k(x,y^{(t-1)})} \pi_F(x) \right] N_{\pi_F}^{(t-1)}[\mathbb{Y}^{(t-1)}]. \quad (1)$$

Here, the initial size  $N^{(0)}$  of the population is given as an initial condition. We omit a strategy  $\pi_F$  or a path  $\mathbb{Y}^{(t)}$  from the notation  $N_{\pi_F}^{(t)}[\mathbb{Y}^{(t)}]$  when they are clear from the context. We can use this dynamical system to define the "fitness" of a strategy  $\pi_F$ . The *cumulative population fitness* of a strategy  $\pi_F$  under  $\mathbb{Y}^{(t)}$  up to time  $t$  is defined as follows:

$$\Lambda^{(t)}(\pi_F | \mathbb{Y}^{(t)}) = \log \frac{N_{\pi_F}^{(t)}[\mathbb{Y}^{(t)}]}{N_{\pi_F}^{(0)}}. \quad (2)$$

The *time-averaged population fitness* of  $\pi_F$  is defined as

$$\lambda(\pi_F) = \lim_{t \rightarrow \infty} \frac{1}{t} \Lambda^{(t)}(\pi_F | \mathbb{Y}^{(t)}), \quad (3)$$

which exists almost surely and independently of  $\mathbb{Y}^{(t)}$  owing to the ergodicity of the environmental state [24,25]. In the following, we call  $\lambda(\pi_F)$  the population fitness in short [26].

When agents learn their strategies [Fig. 1(d)], the number  $N^{(t)}(\pi)$  of the agents with strategy  $\pi$  at time  $t$  becomes as follows:

$$N^{(t)}(\pi) = \left[ \sum_{x \in \mathcal{X}, \pi'} \mathcal{L}(\pi | \pi') e^{k(x,y^{(t-1)})} \pi'(x) \right] N^{(t-1)}(\pi'). \quad (4)$$

Here,  $\mathcal{L}(\pi | \pi')$  is a (possibly stochastic) learning rule, which satisfies  $\sum_{\pi} \mathcal{L}(\pi | \pi') = 1$ . The learning rule  $\mathcal{L}$  can depend on the information available for the agent to learn. We suppose the following sources of available information. Each agent can transmit information to the next generation via epigenetic states or cultures. Specifically, each agent can access the frequency of the types that its ancestors chose. While we do not explicitly consider communications between agents, we will later show in Sec. V that the learning rule without

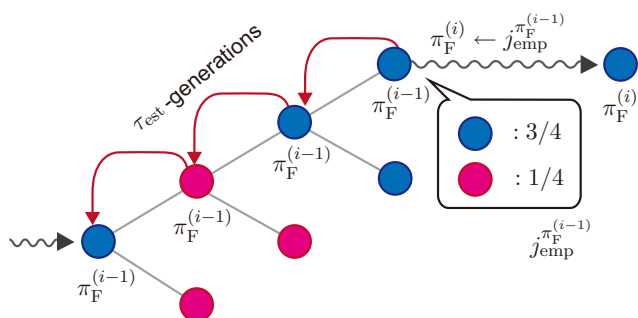


FIG. 2. Schematic representation of ancestral learning. The color of an agent represents its type. After an agent acquires the strategy  $\pi_F^{(i-1)}$  by the  $(i-1)$ th update, its descendants have the same strategy for  $\tau_{\text{est}}$  generations. At the next update, each descendant calculates the empirical distribution  $j_{\text{emp}}^{\pi_F^{(i-1)}}$  and updates the strategy using Eq. (7). This figure corresponds to the case where  $\alpha = 1.0$  and  $\tau_{\text{est}} = 4$ .

communication is sufficient to achieve acceleration of the evolutionary processes by calculating the gradient of population fitness. In addition, we do not assume that the agent can sense the environmental state  $y$ . For further generalization on these assumptions, see Sec. X.

Under this setting, we consider how agents in the population can gradually acquire the optimal strategy:

$$\pi^* = \underset{\pi}{\operatorname{argmax}} \lambda(\pi), \quad (5)$$

by zero- or first-order learning. We note that the optimal strategy is unique because of the concavity of population fitness  $\lambda(\pi)$  with respect to  $\pi$ . The concavity of  $\lambda$  follows from Eq. (15) that we prove in Sec. V.

### III. ANCESTRAL LEARNING

We introduce *ancestral learning* and validate that it can accelerate the evolutionary process. Ancestral learning is self-reinforcement of strategy by a positive feedback. By ancestral learning, an agent updates its strategy every  $\tau_{\text{est}}$  generations, where  $\tau_{\text{est}}$  is a hyperparameter called an *update interval* (see Fig. 2). We suppose that the update occurs at time  $t = i\tau_{\text{est}} - 1$  ( $i = 1, 2, \dots$ ), and the initial strategy  $\pi_F^{(0)}$  is acquired at time  $-1$  by the zeroth update. An agent at time  $(i-1)\tau_{\text{est}} - 1$  acquires the strategy  $\pi_F^{(i-1)}$  by the  $(i-1)$ th update and its descendants at time  $t' [(i-1)\tau_{\text{est}} \leq t' < i\tau_{\text{est}}]$  have the same strategy. At time  $i\tau_{\text{est}} - 1$ , i.e., at the next update, each descendant calculates the empirical distribution  $j_{\text{emp}}^{\pi_F^{(i-1)}}$  of the ancestor's types back to time  $(i-1)\tau_{\text{est}}$ . Specifically,

$$j_{\text{emp}}^{\pi_F^{(i-1)}}(x) := \frac{1}{\tau_{\text{est}}} \sum_{t'=(i-1)\tau_{\text{est}}}^{i\tau_{\text{est}}-1} \delta_{x,x^{(t')}}, \quad (6)$$

where  $\delta_{x,x^{(t' )}}$  is Kronecker's delta, and  $x^{(t' )}$  is the type of the ancestor at time  $t'$ . When  $\pi_F^{(i-1)}$  is clear from the context, we omit it. After obtaining the empirical distribution, the agent updates its strategy using the following rule:

$$\pi_F^{(i)} \leftarrow (1 - \alpha)\pi_F^{(i-1)} + \alpha j_{\text{emp}}^{\pi_F^{(i-1)}}, \quad (7)$$

where  $\alpha$  is a hyperparameter called a *learning rate*. In this rule of ancestral learning, the strategy after the  $i$ th update,  $\pi_F^{(i)}$ , is a mixture of the previous strategy  $\pi_F^{(i-1)}$  with the frequency  $j_{\text{emp}}^{\pi_F^{(i-1)}}$  of types chosen by its ancestors. If the learning rate is close to 1, that is,  $\alpha \approx 1$ , the updated strategy  $\pi_F^{(i)}$  becomes identical to the ancestor's type frequency. If  $\alpha$  is small, the information of the ancestor's type is gradually assimilated into the strategy. Ancestral learning coincides with Xue's rule [7] when  $\tau_{\text{est}} = 1$ . In addition, the rule does not require communication between agents.

Ancestral learning is a biologically reasonable learning rule. The only information used in the rule is the empirical distribution  $j_{\text{emp}}$  of the ancestor's types, which can be stored and transmitted via epigenetic states or memes, as discussed in the Introduction. Owing to this property, we call  $j_{\text{emp}}$  ancestral information. In addition, the memory necessary to store  $j_{\text{emp}}$  is reasonably small. In Sec. V, we prove that the compressed information  $j_{\text{emp}}$  instead of the whole path  $\mathbb{X}^{(t)}$  of the ancestor's type is sufficient to attain the optimal strategy. The update rule of ancestral learning seems natural as it is similar to Hebb's rule [27] as pointed out in Ref. [7]. Hebb's rule describes self-reinforcement through positive feedback, so that the synaptic connection between activated and coactivated neurons is strengthened.

The intuitive explanation for why ancestral learning can attain the optimal strategy is that replicating the types of survived ancestors is likely to contribute to the survival of the descendants. Due to the growth competition among the population, the empirical distribution  $j_{\text{emp}}$  of ancestors' types deviates from the strategy  $\pi_F$ , and  $j_{\text{emp}}$  seen as a strategy has a greater population fitness than  $\pi_F$ . This deviation, known as survivorship bias, works as the driving force of ancestral learning.

To see the intuition more precisely, let us consider a simple case where the environment is constant,  $\mathcal{Y} = \{*\}$ , the learning rate  $\alpha = 1.0$ , and the update interval  $\tau_{\text{est}}$  is sufficiently long. In this case, the individual fitness only depends on type, and we can omit  $y$  in  $e^{k(x,y)}$  as  $e^{k(x)}$ . The optimal strategy  $\pi_F^*$  is as follows:

$$\begin{aligned} \pi_F^*(x^*) &= 1, & x^* &= \underset{x \in \mathcal{X}}{\operatorname{argmax}} k(x) \\ \pi_F^*(x) &= 0 & \text{otherwise,} \end{aligned} \quad (8)$$

which means that  $\pi_F^*$  always selects the type  $x^*$  that maximizes the individual fitness  $e^{k(x)}$ . We calculate  $j_{\text{emp}}^{\pi_F^{(i-1)}}$  to observe how ancestral learning updates the strategy  $\pi_F^{(i-1)}$  and to check that  $\pi_F^{(i)}$  converges to the optimal  $\pi_F^*$  as  $i \rightarrow \infty$ . Since  $j_{\text{emp}}(x) = 1/\tau_{\text{est}} (\sum_{t'=(i-1)\tau_{\text{est}}}^{i\tau_{\text{est}}-1} \delta_{x,x^{(t')}})$  is the sum of independent and identically distributed random variables  $\{\delta_{x,x^{(t')}}\}_{t'}$ , the law of large numbers implies the following:

$$j_{\text{emp}}(x) \approx \langle \delta_{x,x^{(t')}} \rangle, \quad (9)$$

when  $\tau_{\text{est}}$  is sufficiently long (we discuss the case when  $\tau_{\text{est}}$  is not large and show that a small learning rate  $\alpha$  can compensate for the small  $\tau_{\text{est}}$  in Sec. IX). We can interpret  $\langle \delta_{x,x^{(t')}} \rangle$  as a probability defined as follows. Recall that an agent at time  $(i-1)\tau_{\text{est}} - 1$  acquires  $\pi_F^{(i-1)}$  via the update by ancestral learning and that its descendants have the same strategy

until the next update at time  $i\tau_{\text{est}} - 1$ . Let us consider the subpopulation consisting of the descendants. We choose an agent at time  $i\tau_{\text{est}}$  from the subpopulation uniformly at random. Under this setting,  $\langle \delta_{x,x^{(t')}} \rangle$  is the probability  $\pi_B(x)$  that the ancestor of the chosen agent at time  $t'$  expresses type  $x$ . When we choose an agent at time  $t' + 1$  whose parent expressed type  $x$ , the expected number of its descendants at time  $i\tau_{\text{est}}$  is independent of  $x$ . We denote this expected value by  $C^{(t'+1)}$ . Let  $N^{(t')}$  be the number of agents in the subpopulation at time  $t'$ . Using these quantities, we can calculate the number of agents at time  $i\tau_{\text{est}}$ , whose parent at time  $t'$  expressed type  $x$  as  $C^{(t'+1)}e^{k(x)}\pi_F^{(i-1)}(x)N^{(t')}$ . Additionally, the total number of agents at time  $i\tau_{\text{est}}$  is  $\sum_{x' \in \mathcal{X}} C^{(t'+1)}e^{k(x')}\pi_F^{(i-1)}(x')N^{(t')}$ . Therefore, the probability  $\pi_B(x)$  is

$$\begin{aligned} \pi_B(x) &= \frac{C^{(t'+1)}e^{k(x)}\pi_F^{(i-1)}(x)N^{(t')}}{\sum_{x' \in \mathcal{X}} C^{(t'+1)}e^{k(x')}\pi_F^{(i-1)}(x')N^{(t')}} \\ &= \frac{e^{k(x)}\pi_F^{(i-1)}(x)}{\sum_{x' \in \mathcal{X}} e^{k(x')}\pi_F^{(i-1)}(x')}. \end{aligned} \quad (10)$$

Equations (10) and (9) indicate that  $j_{\text{emp}} \approx \pi_B$  when  $\tau_{\text{est}}$  is sufficiently large. The probability  $\pi_B$  is called the *retrospective process* of  $\pi_F^{(i-1)}$  for a constant environment [28–31]. The retrospective process is biased so that  $\pi_B(x^*)$ , the probability of switching to the optimal type  $x^*$ , is larger than  $\pi_F^{(i-1)}(x^*)$ ; therefore, it is better fitted to the environment. As ancestral learning updates  $\pi_F^{(i-1)}$  to  $j_{\text{emp}}$  when  $\alpha = 1.0$ , the strategy becomes  $\pi_F^{(i)}(x) \propto e^{ik(x)}\pi_F^{(0)}(x)$  after the  $i$ th update. Consequently,  $\pi_F^{(i)} \rightarrow \pi^*$  as  $i \rightarrow \infty$ .

We next consider the case where the environment is not constant. We calculate  $j_{\text{emp}}^{\pi_F^{(i-1)}}$  as with the constant-environment case. Since the environmental state  $y^{(t)}$  at each time independently follows  $Q(y)$ , the law of large numbers implies that

$$j_{\text{emp}}(x) \approx \langle \langle \delta_{x,x^{(t')}} | y \rangle \rangle_{Q(y)}, \quad (11)$$

where  $\langle \delta_{x,x^{(t')}} | y \rangle$  is the conditional expectation of  $\delta_{x,x^{(t')}}$ , given the environmental state at time  $t'$  is  $y$ . We can interpret  $\langle \delta_{x,x^{(t')}} | y \rangle$  as the conditional probability  $\pi_B(x | y)$  defined as follows. Let us consider an agent that acquires strategy  $\pi_F^{(i-1)}$  at time  $(i-1)\tau_{\text{est}} - 1$  and the subpopulation that consists of its descendants as before. Suppose that the environmental state at time  $t'$  is  $y$ . We choose an agent at time  $i\tau_{\text{est}}$  uniformly at random from the subpopulation. In this setting,  $\langle \delta_{x,x^{(t')}} | y \rangle$  is the probability  $\pi_B(x | y)$  that the ancestor of the chosen agent at time  $t'$  expresses type  $x$ . Using a similar argument, we can prove that

$$\pi_B(x | y) = \frac{e^{k(x,y)}\pi_F(x)}{\sum_{x' \in \mathcal{X}} e^{k(x',y)}\pi_F(x')}. \quad (12)$$

The probability  $\pi_B(x | y)$  is also called a retrospective process and is fitted to the environmental state  $y$  better than  $\pi_F$ . Equations (12) and (11) indicate that  $j_{\text{emp}}(x)$  converges to the averaged retrospective process  $\bar{\pi}_B(x) := \sum_y \pi_B(x | y)Q(y)$ . Therefore,  $\pi_F$  is updated to the mixture of strategies  $\pi_B(x | y)$ , each of which is better fitted to the corresponding environmental state. We numerically (Sec. IV) and theoretically (Sec. V) prove that the update to such a mixed strategy leads to the optimal strategy.

#### IV. ANCESTRAL LEARNING CAN ACCELERATE EVOLUTIONARY PROCESSES

Next, we validate that learning can accelerate evolutionary processes by numerically showing that the optimal type-switching strategy is acquired with ancestral learning faster than with zero-order mutational rules.

We simulated the evolutionary process using a multitype branching process in a random environment [32,33]. We simulated the dynamical system defined by Eq. (4) while considering the individuality and finite size of the population. In the simulation, we set  $\mathcal{X} = \mathcal{Y} = \{0, 1, 2\}$  [Fig. 3(a)]. We also set  $Q(0) = 0.6$  and  $Q(1) = Q(2) = 0.2$ . Each agent with type  $x$  under environmental state  $y$  had four daughters if  $x = y$  and one daughter otherwise. In short,  $e^{k(x,y)} = 4$  if  $x = y$  and  $e^{k(x,y)} = 1$  otherwise. We represent the strategy  $\pi_F$  as a vector of the form  $(\pi_F(0), \pi_F(1), \pi_F(2))$ . Owing to the symmetry of  $e^{k(x,y)}$ , the zeroth component  $\pi^*(0)$  of the optimal strategy is higher than that of the others. We started the simulation from a single agent, whose initial strategy is  $\pi_F^{(0)} = (1/3, 1/3, 1/3)$ . We limited the number of agents in the population to  $N_{\text{max}} = 30$  to avoid the intractability of the numerical experiment owing to the exponential growth of the number of agents. If the number of agents in the next generation exceeded  $N_{\text{max}}$ , we selected  $N_{\text{max}}$  agents uniformly at random.

We investigated three learning rules. Each learning rule updates the strategy at every time step, that is,  $\tau_{\text{est}} = 1$ . The first learning rule was ancestral learning with the learning rate  $\alpha = 0.01$ . The second and the third were the zero-order mutational rules. As there are innumerable zero-order learning rules, we chose two representatives to perform the control experiments for ancestral learning. The second learning rule was  $\pi_F' \leftarrow (1 - \alpha)\pi_F + \alpha\delta_{x,x_{\text{rand}}}$ , where  $\pi_F$  and  $\pi_F'$  were the strategies of the agent before and after the update, respectively, and  $x_{\text{rand}}$  was chosen uniformly at random from  $\mathcal{X}$ . In biological systems, this rule can be seen as a random mutation of  $\pi_F$ , whose rate is constant. The trajectory of  $\pi_F$  updated by this rule is a random walk over  $\mathbb{R}^{\mathcal{X}}$  if no growth occurs, that is,  $e^{k(x,y)} = 1$  for all  $x \in \mathcal{X}$  and  $y \in \mathcal{Y}$ . We therefore call this learning rule a random walk. The third learning rule was  $\pi_F' \leftarrow (1 - \alpha)\pi_F + \alpha\delta_{x,x_{\text{samp}}}$ , where  $x_{\text{samp}}$  was sampled from the discrete distribution  $\pi_F$ . In biological systems, this rule can be seen as a mutation of  $\pi_F$ , whose rate is dependent on the current  $\pi_F$ . The change in mutation rate is known as adaptive mutation [34]. Therefore, we call this learning rule an adaptive random walk. The adaptive random walk coincides with ancestral learning if no growth occurs. In this sense, the adaptive random walk is a control for observing the effect of population growth on ancestral learning.

Figure 3 is the simulation of the three learning rules, which shows that ancestral learning accelerates the evolutionary process. We showed lineage trees up to  $t = 50$ . The population fitness with ancestral learning increased faster than with the other learning rules [Fig. 3(b)] along the lineage of the most successful agent, whose population fitness was the maximum of the agents at the end of each lineage tree. Acceleration of the evolutionary process was also observed at the lineage tree level [Figs. 3(d)–(f)]. In Figs. 3(g)–(i), we selected the lineage of the most successful agent in each lineage tree and plotted the trajectory of  $\pi_F$  along the lineage. In the population

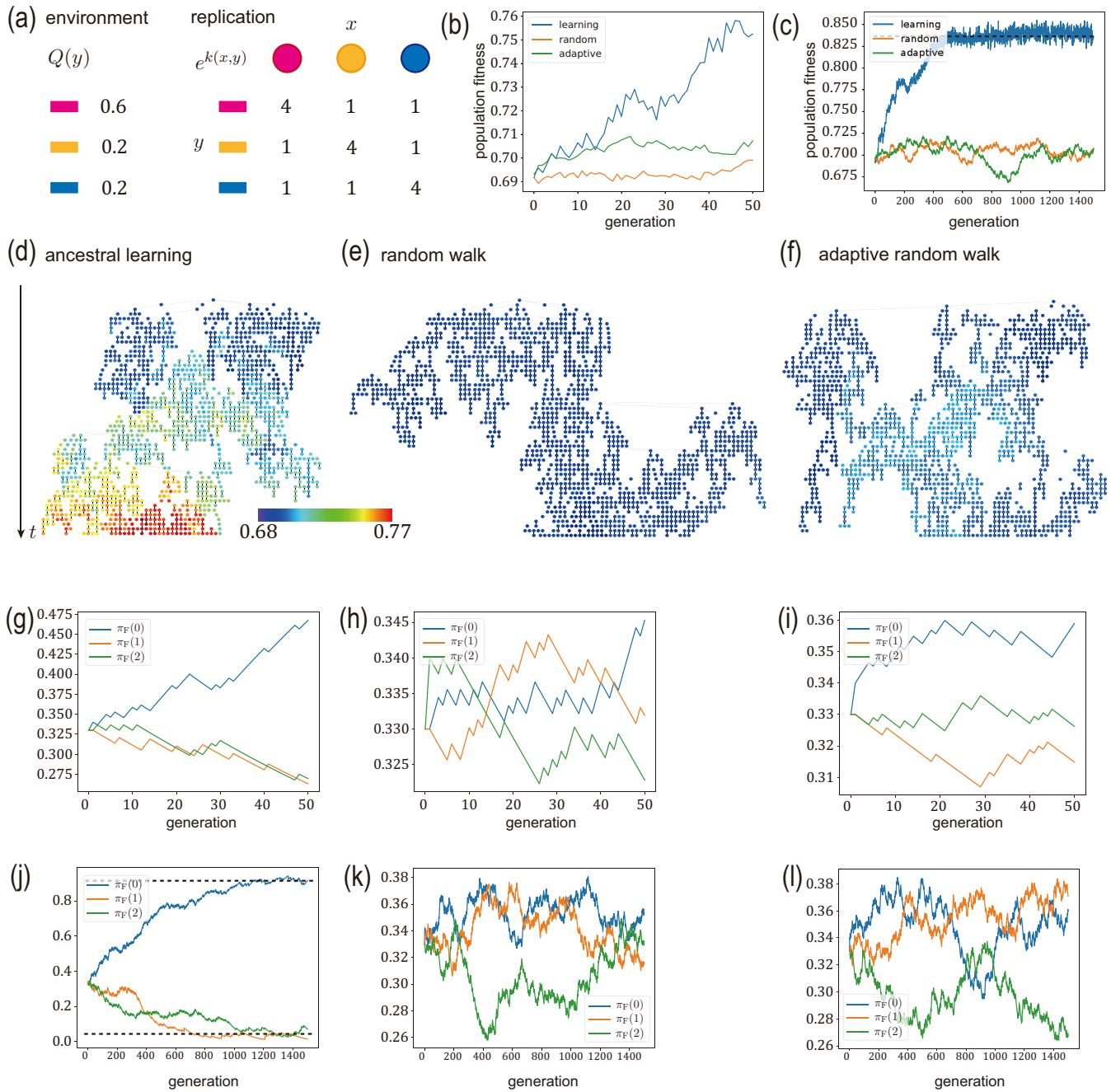


FIG. 3. Numerical experiments to investigate ancestral learning. (a) The parameters of the model. In the panel,  $0 \in \mathcal{Y}$  corresponds to red, 1 to yellow, and 2 to blue. The red environmental state occurs more frequently than the others. Each agent with type  $x$  under environmental state  $y$  has four daughters if  $x = y$  and one daughter otherwise. (b) Population fitness trajectories until  $t = 50$  for the lineages of the most successful agents, whose  $\lambda$  was the maximum of the agents at the end of each lineage tree. Each curve corresponds to an agent that adopts a different learning rule. The blue, orange, and green curves represented ancestral learning, the random walk, and the adaptive random walk, respectively. Ancestral learning resulted in the best increase in population fitness among the three learning rules. (c) The same plot as in (b) until  $t = 1500$ . The dotted line showed the population fitness for the most successful agent with ancestral learning at  $t = 1500$ . In the longer simulation, we could see the convergence of population fitness in the population with ancestral learning. (d)–(f) Simulated lineage trees for the agents that (d) adopt ancestral learning, (e) the random walk, and (f) the adaptive random walk. Each point corresponded to an agent, and its color represents the population fitness  $\lambda$  of the agent. Black lines connected parents to their daughters. (g)–(i) The trajectories of strategy  $\pi_F$  until  $t = 50$  for the lineage of the most successful agent updated by (g) ancestral learning, (h) the random walk, and (i) the adaptive random walk, respectively. (j)–(l) The same plot as (g)–(i) until  $t = 1500$ . In (j), the upper dotted line indicates  $\pi_F(0)$  of the most successful agent, and the lower indicates the average of  $\pi_F(1)$  and  $\pi_F(2)$ . We could see that the strategy of the most successful agent with ancestral learning converged to (0.92, 0.04, 0.04) approximately whereas those with the other learning rules did not converge. The converged strategy was close to the optimal because it satisfied the optimality condition (Theorem 16.2.1 of Ref. [35]) with a small error.

with ancestral learning,  $\pi_F$  moved faster toward the optimal solution [approximately (0.92, 0.04, 0.04)]; see the next paragraph].

To determine whether the optimal strategy was acquired by ancestral learning, we ran another simulation until  $t = 1500$  [Figs. 3(c) and 3(j)–3(l)]. We first checked that the strategy converged; that is, the strategy before and after the update was almost identical when  $t$  is sufficiently large. Then, we verified that the converged strategy is the optimal strategy. We checked the convergence of the strategy along the lineage of the most successful agent with ancestral learning. The strategy converged as the population fitness along the lineage reached a ceiling [Fig. 3(c)]. The convergence was also supported by the trajectory of the strategy [Figs. 3(j)–3(l)]. The converged strategy [approximately (0.92, 0.04, 0.04)] of the most successful agent with ancestral learning was close to the optimal as it satisfied the optimality (Karush-Kuhn-Tucker) condition with a small error (see Theorem 16.2.1 of Ref. [35]).

From these results, we conclude that ancestral learning accelerates the evolutionary process. Since ancestral learning does not use the information from communications among the agents at the same time, these results show that learning can accelerate the evolutionary process even without communication.

## V. ANCESTRAL INFORMATION IS SUFFICIENT TO ESTIMATE GRADIENT

The second problem, whether an agent can estimate the gradient of population fitness from the accessible information, is then addressed. While we have numerically shown that ancestral learning accelerates evolutionary processes, the relationship between ancestral learning and fitness gradient is still unclear. The ancestral information  $j_{\text{emp}}$  used in ancestral learning might be insufficient for estimating the gradient, and communication between the agents in the same generation might be required for the fitness gradient. In this section, we prove that the ancestral information  $j_{\text{emp}}$  is sufficient for estimating the gradient. Theoretically, it implies that an agent can estimate the gradient without communication among agents. This also implies that ancestral learning updates the strategy in the direction of the gradient.

To calculate the gradient of the population fitness, we employ a pathwise formulation and variational principle [31] of the population dynamics. Let us consider the case in which the path of the environmental state is  $\mathbb{Y}^{(t)}$ , and the agents do not learn and stick to a fixed strategy  $\pi_F$ . In this setting, let us calculate the number  $N_{\pi_F}[\mathbb{X}^{(t)} | \mathbb{Y}^{(t)}]$  of the agents at time  $t$  whose path of the type of the ancestors is  $\mathbb{X}^{(t)}$ . By applying Eq. (1) recursively, we have

$$N_{\pi_F}[\mathbb{X}^{(t)} | \mathbb{Y}^{(t)}] = e^{k(\mathbb{X}^{(t)}, \mathbb{Y}^{(t)})} \mathbb{P}_F[\mathbb{X}^{(t)}] N^{(0)}, \quad (13)$$

where the quantities  $k(\mathbb{X}^{(t)}, \mathbb{Y}^{(t)}) := \sum_{t'=0}^{t-1} k(x^{(t')}, y^{(t')})$  and  $\mathbb{P}_F[\mathbb{X}^{(t)}] = \prod_{t'=0}^{t-1} \pi_F(x^{(t')})$  are the pathwise (historical) individual fitness and pathwise forward probability, respectively. With the pathwise formulation, we can represent the cumulative population fitness as

$$\Lambda^{(t)}(\pi_F | \mathbb{Y}^{(t)}) = \log \sum_{\mathbb{X}^{(t)}} e^{k(\mathbb{X}^{(t)}, \mathbb{Y}^{(t)})} \mathbb{P}_F[\mathbb{X}^{(t)}]. \quad (14)$$

Since each  $y^{(t)}$  follows  $Q(y)$  independently, the population fitness satisfies (cf. Refs. [24,25])

$$\lambda(\pi_F) = \langle \log \langle e^{k(x,y)} \rangle_{\pi_F(x)} \rangle_{Q(y)}. \quad (15)$$

The form of  $\log \langle \cdot \rangle_{\pi_F}$  on the right-hand side is equivalent to the scaled cumulant generating function [36], and the following variational principle holds:

$$\lambda(\pi_F) = \left\langle \max_{\pi} \left\{ \sum_{x \in \mathcal{X}} k(x, y) \pi(x) - \mathcal{D}[\pi \| \pi_F] \right\} \right\rangle_{Q(y)}, \quad (16)$$

where  $\pi$  runs over all distributions on  $\mathcal{X}$  and  $\mathcal{D}[\cdot \| \cdot]$  is the Kullback-Leibler (KL) divergence defined by

$$\mathcal{D}[\pi \| \pi'] := \sum_{x \in \mathcal{X}} \pi(x) \log \frac{\pi(x)}{\pi'(x)}. \quad (17)$$

See Appendix A for the proof. By a direct calculation, we can see that the maximizer is  $\pi_B(x | y)$ . We can calculate the derivative of the population fitness from the variational principle:

$$\frac{\partial \lambda(\pi_F)}{\partial \pi_F(x)} = \frac{\bar{\pi}_B(x)}{\pi_F(x)}. \quad (18)$$

See Appendix B for the proof.

We now have all the ingredients to calculate the gradient of the population fitness. As the strategy  $\pi_F$  has a constraint  $\sum_{x \in \mathcal{X}} \pi_F(x) = 1$ , we consider the following definition of the gradient. The *gradient* at  $\pi_F$  under the constraint  $\sum_{x \in \mathcal{X}} \pi(x) = 1$  is defined as

$$\lim_{\epsilon \rightarrow +0} \operatorname{argmax}_{\substack{\delta\pi \\ \pi_F + \delta\pi \in D_\epsilon(\pi_F)}} \{\lambda(\pi_F + \delta\pi)\}, \quad (19)$$

where the limit is one sided from the positive real numbers,  $\delta\pi \in \mathbb{R}^{\mathcal{X}}$  with  $\sum_{x \in \mathcal{X}} (\pi_F(x) + \delta\pi(x)) = 1$ , that is,  $\sum_{x \in \mathcal{X}} \delta\pi(x) = 0$ , and  $D_\epsilon(\pi_F)$  is the sphere around  $\pi_F$  with a radius of  $\epsilon$ . To define the sphere, we use the KL divergence as a natural distance over distributions on  $\mathcal{X}$  as  $D_\epsilon(\pi_F) := \{\pi | \mathcal{D}[\pi \| \pi_F] < \epsilon\}$ . Intuitively, the gradient is the direction in which the population fitness increases the most among all strategies that satisfy the constraint and have the same infinitesimal distance from  $\pi_F$ . The definition is related to a *proximal operator* [37] and coincides with the usual gradient if no constraint is imposed, and if the sphere is defined by the Euclidean distance. We prove that the gradient is directed toward  $\bar{\pi}_B$ , i.e.,

$$\lim_{\epsilon \rightarrow +0} \operatorname{argmax}_{\substack{\delta\pi \\ \pi_F + \delta\pi \in D_\epsilon(\pi_F)}} \{\lambda(\pi_F + \delta\pi)\} \propto \bar{\pi}_B - \pi_F. \quad (20)$$

See Appendix B for the proof.

The results address the second problem. To estimate the gradient, an agent must estimate  $\bar{\pi}_B$ . Based on the discussion in the last paragraph of Sec. III, the ancestral information  $j_{\text{emp}}$  is an unbiased estimator of  $\bar{\pi}_B$ ; that is,  $\langle j_{\text{emp}} \rangle = \bar{\pi}_B$ . Therefore, an agent can estimate the gradient from ancestral information without communication among agents in the same generation. The explicit formula of the gradient also indicates that ancestral learning updates the strategy in the direction of the gradient. The direction  $\pi_F^{(i+1)} - \pi_F^{(i)}$  of the

update of the strategy by ancestral learning equals the right-hand side of Eq. (20). In particular, ancestral learning finds the optimal strategy if the learning rate is sufficiently small because  $\lambda$  is concave. In addition, while we assumed here that  $\tau_{\text{est}}$  is infinitely large, we also show in Sec. IX that this assumption can be relaxed approximately for small  $\tau_{\text{est}}$  by choosing the learning rate  $\alpha$  sufficiently small.

## VI. FISHER'S FUNDAMENTAL THEOREM FOR ANCESTRAL LEARNING

We address the last problem, the quantification of the acceleration of evolutionary processes by learning, by extending the FF-thm to ancestral learning. Ancestral learning may increase the population fitness much faster in certain environments than others, depending on the stochastic property  $Q(y)$  of the environment. In addition, the acceleration can also depend on the update interval  $\tau_{\text{est}}$  and the learning rate  $\alpha$ . The extension of the conventional FF-thm to ancestral learning will improve our understanding of the dependency of acceleration on  $Q(y)$ , as well as when and why learning becomes beneficial.

We will first review the conventional FF-thm for natural selection [22]. The FF-thm relates to the speed of evolution and the variance of individual fitness in the population. To illustrate this, we consider the following fixed-type population dynamics in a constant environment: The set of types is  $\mathcal{X}$  as Eq. (1), the daughter type is the same as that of the parent, and the environment is constant,  $\mathcal{Y} = \{*\}$ . The individual fitness of type  $x$  is  $e^{k(x)}$ . Here, we omit the dependency of individual fitness  $e^{k(x,*)}$  on the environmental state  $*$  as in the derivation of Eq. (10). Under this setting, the number  $N^{(t)}(x)$  of the agent with type  $x$  at time  $t$  is as follows:

$$N^{(t)}(x) = e^{k(x)}N^{(t-1)}(x). \quad (21)$$

As we are interested in statistics of the population, such as the variance of the individual fitness, we focus on the fraction  $p^{(t)}(x) := N^{(t)}(x)/\sum_{x \in \mathcal{X}} N^{(t)}(x)$  of the agents with type  $x$  at time  $t$  instead of  $N^{(t)}(x)$ . The time evolution of  $p^{(t)}$  was derived from Eq. (21) and is expressed as

$$p^{(t)}(x) = \frac{e^{k(x)}p^{(t-1)}(x)}{\sum_{x' \in \mathcal{X}} e^{k(x')}p^{(t-1)}(x')}. \quad (22)$$

We define the *covariance* of the random variables  $f(x)$  and  $g(x)$  with respect to a probability distribution  $p(x)$  over  $\mathcal{X}$  as

$$\text{Cov}_p[f(x), g(x)] := \langle f(x)g(x) \rangle_p - \langle f(x) \rangle_p \langle g(x) \rangle_p. \quad (23)$$

From this, the *variance* is also defined as

$$\mathbb{V}_p[f(x)] := \text{Cov}_p[f(x), f(x)]. \quad (24)$$

One of the measures of evolutionary speed is the gain of mean individual fitness  $\langle e^{k(x)} \rangle_{p^{(t)}}$ . The gain satisfies the following relationship owing to Eq. (22):

$$\Delta \langle e^{k(x)} \rangle_{p^{(t)}} := \langle e^{k(x)} \rangle_{p^{(t)}} - \langle e^{k(x)} \rangle_{p^{(t-1)}} \quad (25)$$

$$= \mathbb{V}_{p^{(t-1)}}[e^{k(x)}] / \langle e^{k(x)} \rangle_{p^{(t-1)}}. \quad (26)$$

See Appendix C for the detailed proof. The equation reveals the relationship between evolutionary speed and the variance of individual fitness in the population. This equation is called

the FF-thm for natural selection [38]. As we are not interested in the mean individual fitness but the population fitness defined by

$$\lambda^{(t)} := \log \frac{\sum_{x \in \mathcal{X}} N^{(t)}(x)}{\sum_{x \in \mathcal{X}} N^{(t-1)}(x)}, \quad (27)$$

at time  $t$ , we derive an FF-thm for population fitness. We define variants of the covariance and variance, which we call *log covariance* and *log variance*, to extend the conventional FF-thm by

$$\log\text{-Cov}_p[f(x), g(x)] := \log \frac{\langle f(x)g(x) \rangle_p}{\langle f(x) \rangle_p \langle g(x) \rangle_p}, \quad (28)$$

$$\log\text{-}\mathbb{V}_p[f(x)] := \log\text{-Cov}_p[f(x), f(x)], \quad (29)$$

respectively. The log covariance measures the similarity of two random variables as the covariance does because the log covariance monotonically increases with respect to the covariance. Indeed, we can prove

$$\log\text{-Cov}_p[f(x), g(x)] = \log \left( 1 + \frac{\text{Cov}_p[f(x), g(x)]}{\langle f(x) \rangle_p \langle g(x) \rangle_p} \right) \quad (30)$$

by direct calculation. By using these quantities, we can obtain an extended FF-thm for population fitness using a similar argument as in Eq. (26),

$$\Delta \lambda^{(t)} := \lambda^{(t)} - \lambda^{(t-1)} = \log\text{-}\mathbb{V}_{p^{(t-1)}}[e^{k(x)}]. \quad (31)$$

See Appendix C for the proof. This equation reveals the relationship between the speed of evolutionary process measured by the gain of the population fitness and the log variance of the individual fitness in the population.

The FF-thm for population fitness is closely related to ancestral learning. To see this, we first consider a simple case in which the environment is constant,  $\mathcal{Y} = \{*\}$ , the learning rate  $\alpha = 1.0$ , and  $\tau_{\text{est}} \approx \infty$ . Under this setting, we measure the acceleration of the evolutionary process by ancestral learning using  $\Delta \lambda^{(i)} := \lambda(\pi_{\text{F}}^{(i)}) - \lambda(\pi_{\text{F}}^{(i-1)})$ , where  $\pi_{\text{F}}^{(i-1)}$  and  $\pi_{\text{F}}^{(i)}$  are the strategies of an agent before and after the update by ancestral learning. The gain  $\Delta \lambda^{(i)}$  of the population fitness depends on ancestral learning and is independent of natural selection. We can therefore regard  $\Delta \lambda^{(i)}$  as a measure of the acceleration. By using these preparations, we find that the update by ancestral learning and gain  $\Delta \lambda^{(i)}$  have a close connection with the FF-thm for population fitness. In Sec. III, we have shown that the update of ancestral learning is  $\pi_{\text{F}}^{(i)} \leftarrow \pi_{\text{B}}^{(i-1)}$ , where  $\pi_{\text{B}}^{(i-1)}$  is the retrospective process of  $\pi_{\text{F}}^{(i-1)}$ . This update is equivalent to that in Eq. (22) if we identify  $p^{(t)}$  with  $\pi_{\text{F}}^{(i)}$ . In addition, the gain  $\Delta \lambda^{(i)}$  is equivalent to the left-hand side of Eq. (31) if we identify  $p^{(t)}$  with  $\pi_{\text{F}}^{(i)}$  again. Owing to these two equivalences, we can obtain an FF-thm for ancestral learning by substituting  $p^{(t)}$  with  $\pi_{\text{F}}^{(i)}$  in Eq. (31):

$$\Delta \lambda^{(i)} := \lambda(\pi_{\text{F}}^{(i)}) - \lambda(\pi_{\text{F}}^{(i-1)}) = \log\text{-}\mathbb{V}_{\pi_{\text{F}}^{(i-1)}}[e^{k(x)}]. \quad (32)$$

This theorem reveals the relationship between gains in population fitness by using ancestral learning and the log variance of individual fitness of the strategy.

The theorem also reveals the trade-off between the acceleration  $\Delta \lambda^{(i)}$  and the population fitness  $\lambda(\pi_{\text{F}}^{(i)})$  by showing



that the acceleration is large when the agent expresses a variety of types. An interpretation is that the agent can obtain information about which type is best fitted to the environment by expressing a variety of types. We call such a situation exploratory. In contrast, an agent with the optimal strategy always expresses the same type in this setting [Eq. (8)]. Therefore, the theorem implies that the acceleration is almost zero when the strategy is close to the optimal value, and  $\lambda(\pi_F^{(i)})$  is large. We call such a situation exploitative. Thus, we can see the so-called exploration-exploitation trade-off in this setting.

We can further extend the FF-thm for ancestral learning to cases where the environment is not constant, as follows:

$$\Delta\lambda^{(i)} = \langle \log\text{-Cov}_{\pi_F^{(i-1)}}[e^{k(x,y)}, e^{k(x,y')}]\rangle_{Q(y)Q(y')} + \mathcal{D}[Q(y)Q(y')\|\bar{Q}^{(i)}(y' | y)Q(y)], \quad (33)$$

where

$$\bar{Q}^{(i)}(y' | y) \propto \sum_{x \in \mathcal{X}} e^{k(x,y)} \pi_B^{(i-1)}(x | y') Q(y'). \quad (34)$$

See Appendix D for the proof.

Note that the above equation can be reduced to Eq. (32) if the environment is constant. Equation (33) is different from the FF-thm of natural selection in a nonconstant environment, as the time for the evolution of  $p^{(t)}$  is different from the update for  $\pi_F^{(i)} \leftarrow \bar{\pi}_B^{(i-1)}$  with ancestral learning. The time for the evolution of  $p^{(t)}$  in a nonconstant environment is stochastic and governed by the following:

$$p^{(t)}(x) = \frac{e^{k(x,y)} p^{(t-1)}(x)}{\sum_{x' \in \mathcal{X}} e^{k(x',y)} p^{(t-1)}(x')}, \quad (35)$$

with probability  $Q(y)$ .

## VII. MEASURES TO CHARACTERIZE ANCESTRAL LEARNING

Using the terms that appear in Eq. (33), we can quantitatively characterize different aspects of the strategies during and after learning. We define an actual gain  $\Delta_{ac}\lambda^{(i)}$  and an expected gain  $\Delta_{ex}\lambda^{(i)}$  using the left- and right-hand sides of Eq. (33):

$$\Delta_{ac}\lambda^{(i)} := \lambda(\pi_F^{(i)}) - \lambda(\pi_F^{(i-1)}) \quad (36)$$

and

$$\Delta_{ex}\lambda^{(i)} := \tilde{\Sigma}^{(i)} + \text{KL}^{(i)}, \quad (37)$$

where  $\tilde{\Sigma}^{(i)}$  and  $\text{KL}^{(i)}$  are variance and KL terms of  $\Delta_{ex}\lambda^{(i)}$ , respectively, defined as

$$\tilde{\Sigma}^{(i)} := \langle \log\text{-Cov}_{\pi_F^{(i-1)}}[e^{k(x,y)}, e^{k(x,y')}]\rangle_{Q(y)Q(y')} \quad (38)$$

and

$$\text{KL}^{(i)} := \mathcal{D}[Q(y)Q(y')\|\bar{Q}^{(i)}(y' | y)Q(y)]. \quad (39)$$

The reason for the additional KL term [Eq. (39)] appearing in Eq. (33) is attributed to the existence of two representative strategies: *specialist* and *generalist*. Each term [Eqs. (38) and (39)] of the expected gain [Eq. (37)] is associated with one of the representative strategies and corresponds to the gain of the population fitness by acquiring the corresponding strategy.

A specialist strategy is defined as a situation in which an agent expresses a small subset of types that are fitted to the environment. Formally, a strategy is specialized to  $\mathcal{X}' \subsetneq \mathcal{X}$  if  $\pi_F(x) > 0$  for  $x \in \mathcal{X}'$  and  $\pi_F(x) = 0$  otherwise. An example is the optimal strategy [Eq. (8)] for a constant environment which is specialized to the single optimal type  $\{x^*\}$ . The specialist strategy is beneficial when the environment is constant or the environmental states  $y \in \mathcal{Y}$  are similar to each other, because an agent can survive by expressing not all but a few types in such situations. Here, the similarity between the environmental states  $y$  and  $y'$  is defined as the closeness of  $e^{k(x,y)}$  and  $e^{k(x,y')}$  for all  $x \in \mathcal{X}$  (see the next paragraph for the formal definition). However, if the environmental states are dissimilar, an agent cannot produce descendants efficiently by specializing to only a few types, as these types are not adaptive to some environmental states. An agent should stochastically choose types from a variety of alternatives to reduce the risk of specialization. The probability of expressing a type should be determined such that the strategy has a greater population fitness. Even if the strategy is specialized to a subset  $\mathcal{X}'$  with  $\#\mathcal{X}' > 1$ , the probability  $\pi_F(x)$  for  $x \in \mathcal{X}'$  should be determined to maximize  $\lambda$ . We define the generalist strategy to  $\mathcal{X}'$  as the stochastic expression of the types in  $\mathcal{X}'$ , whose probabilities are positive and are set such that the population fitness is maximized. In general, the optimal strategy is a combination of specialist and generalist. For example, let us examine the optimal strategy  $\pi_F^* = (0.72, 0.0, 0.28)$  in the model shown in Fig. 4(j), which is calculated numerically. The strategy is specialized to  $\mathcal{X}' = \{0, 2\}$  but of generalist in  $\mathcal{X}'$ .

During the evolutionary process with learning, an agent attains the optimal strategy by acquiring two representative strategies. The variance and KL terms of the expected gain,  $\tilde{\Sigma}^{(i)}$  and  $\text{KL}^{(i)}$ , correspond to the gains in population fitness by acquiring the respective strategy. The variance term  $\tilde{\Sigma}^{(i)}$  measures the gain of population fitness by acquiring the specialist strategy, whereas the KL term  $\text{KL}^{(i)}$  does so by acquiring the generalist strategy. To see this interpretation, we rewrite an updated strategy  $\pi_F^{(i)}$ . We proved that  $\pi_F^{(i)} = \bar{\pi}_B^{(i-1)}$  in Sec. III when  $\tau_{\text{est}} \approx \infty$ . By definition,

$$\bar{\pi}_B^{(i-1)}(x) = \left\langle \frac{e^{k(x,y)}}{\langle e^{k(x',y')} \rangle_{\pi_F^{(i-1)}(x')}} \right\rangle_{Q(y)} \pi_F^{(i-1)}(x) \propto \langle e^{k(x,y)} \rangle_{Q(y)} \pi_F^{(i-1)}(x). \quad (40)$$

This equation transforms the probability distribution  $\pi_F^{(i-1)}$  into  $\bar{\pi}_B^{(i-1)}$  by multiplying  $\langle e^{k(x,y)} \rangle_{Q(y)}$  for each  $x \in \mathcal{X}$ . In the transformation, the normalization factor is  $\langle e^{k(x',y')} \rangle_{\pi_F^{(i-1)}(x')Q(y)}$ . Let us examine the multiplicative factors. For convenience, we define a vector  $(\langle e^{k(x,y)} \rangle_{Q(y)})_{x \in \mathcal{X}} \in \mathbb{R}^{\mathcal{X}}$  by collecting the multiplicative factors for  $x \in \mathcal{X}$ . It is the average of the vectors  $F_y := (e^{k(x,y)})_{x \in \mathcal{X}} \in \mathbb{R}^{\mathcal{X}}$  defined for each  $y$ . We regard  $F_y$  as a representation of the environmental state  $y$  by embedding it into  $\mathbb{R}^{\mathcal{X}}$  [Figs. 4(e) and 4(h)]. The embedding can be used to measure the similarity between environmental states  $y$  and  $y'$  using  $\log\text{-Cov}_{\pi_F^{(i)}}[F_y(x), F_{y'}(x)]$ . By considering the normalization factor  $\langle e^{k(x',y')} \rangle_{\pi_F^{(i-1)}(x')Q(y)}$ , we also define a scaled

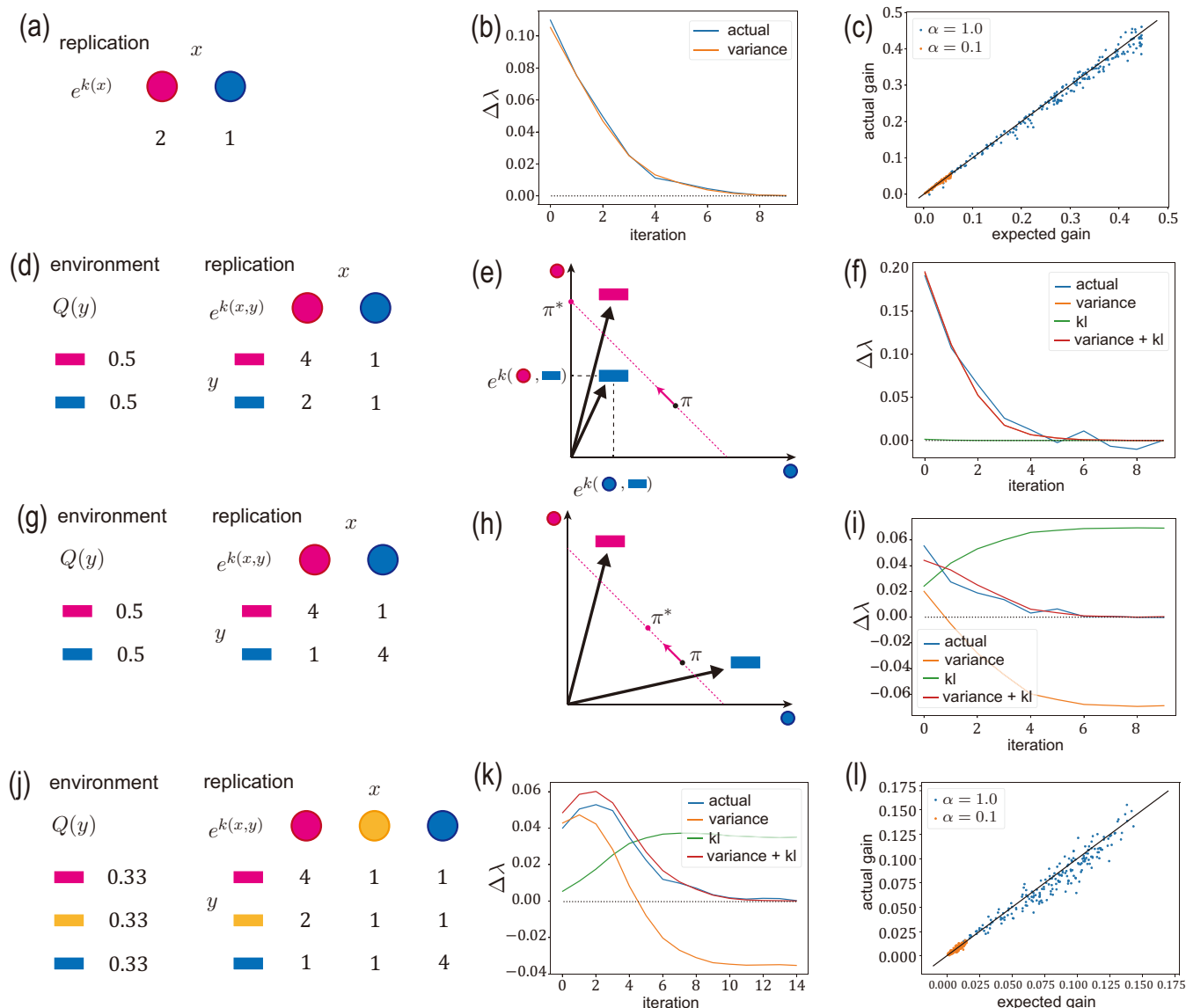


FIG. 4. Numerical validation of FF-thm for ancestral learning. The learning rate is  $\alpha = 1.0$  unless otherwise specified. (a)–(c) Constant environment model. (a) The parameters of the model. (b) Trajectories of the actual gain  $\Delta_{ac}\lambda^{(i)}$  and the variance term  $\tilde{\Sigma}^{(i)}$  [Eqs. (36) and (38)] along the lineage of an agent. Notice that  $\tilde{\Sigma}^{(i)} = \Delta_{ex}\lambda^{(i)}$  [Eq. (37)] when the environment is constant. The dotted black line represents  $\Delta\lambda = 0$ . At each update, we observe  $\Delta_{ac}\lambda^{(i)} \approx \tilde{\Sigma}^{(i)}$ . (c) Comparison between  $\Delta_{ac}\lambda^{(1)}$  and  $\tilde{\Sigma}^{(1)}$  when an agent had a randomly generated initial strategy. For most of the strategies,  $\Delta_{ac}\lambda^{(1)} \approx \tilde{\Sigma}^{(1)}$  was observed when the learning rate is  $\alpha = 1.0$  or  $\alpha = 0.1$ . (d)–(f) Similar environment model. (d) The parameters of the model. (e) Illustration of the representation of environmental state  $y$  in  $\mathbb{R}^X$  using the embedding vector  $F_y$ . The environmental state  $y$  is represented such that the  $x$ th component of  $F_y$  is  $e^{k(x,y)}$ . The environmental states are similar since these two embedded vectors point to similar directions. The optimal solution  $\pi^*$  in this model is specialist. Geometrically, the strategy lies on the red dotted line  $\sum_{x \in \mathcal{X}} \pi_F(x) = 1$ , and the optimal is on the axis corresponding to the red type. Thus, the strategy moved toward  $\pi^*$  on this line. (f) Comparison between  $\Delta_{ac}\lambda^{(i)}$ ,  $\tilde{\Sigma}^{(i)}$ ,  $KL^{(i)}$  [Eq. (39)], and  $\Delta_{ex}\lambda^{(i)}$ . As the FF-thm for nonconstant environments [Eq. (33)] has an additional KL term, the KL term  $KL^{(i)}$  and the expected gain  $\Delta_{ex}\lambda^{(i)}$  were shown in addition to (b). We could observe that the FF-thm holds and the variance term dominated. (g)–(i) Dissimilar environment model. (g) The parameters of the model. (h) An illustration of the embedding of environmental states in this model. From this embedding, we can see that the environmental states are dissimilar. In this case, the optimal strategy is generalist and should lie in the middle of the red dotted line. (i) The same plot as (f) in this model. We observed that the FF-thm holds and the KL term was not negligible. At last,  $|\tilde{\Sigma}^{(i)}| \approx |KL^{(i)}|$  was achieved. (j)–(l) Combined model. (j) The parameters of the model. (k) The same plot as (b) with the additional trajectories of the KL term  $KL^{(i)}$  and the expected gain  $\Delta_{ex}\lambda^{(i)}$ . Since the variance term dropped earlier than the KL term, we can see that an agent learned a specialist strategy first and then acquired a generalist strategy. (l) The same plot as (c) in this model. For most of the strategies, we could see that  $\Delta_{ac}\lambda^{(1)} \approx \Delta_{ex}\lambda^{(1)}$  holds when the learning rate is  $\alpha = 1.0$  or  $\alpha = 0.1$ .

embedding  $f_y$  as follows:

$$f_y(x) := \frac{F_y(x)}{\langle e^{k(x,y)} \rangle_{\pi_F^{(i-1)}(x')}} \quad (41)$$

which depends on the current strategy  $\pi_F^{(i-1)}$  in addition to  $y$ . We use scaled embedding to rewrite Eq. (40) as

$$\bar{\pi}_B^{(i-1)}(x) = \langle f_y(x) \rangle_{Q(y)} \pi_F^{(i-1)}(x). \quad (42)$$

The updated strategy  $\bar{\pi}_B^{(i-1)}$  is closer to the specialist strategy when environmental states are more similar since if each  $f_y$  has similar peaks (larger components), so is their average  $\bar{\pi}_B^{(i-1)}$  [Fig. 4(e)]. The iteration of such an update leads to specialization to the types where the peaks lie on. We will later see in the next paragraph that the variance term [Eq. (38)] measures the similarity of environmental states and corresponds to the gain of the population fitness by being of specialist strategy. In contrast,  $\bar{\pi}_B^{(i-1)}$  is closer to a generalist strategy when the environmental states are more dissimilar as if each  $f_y$  has different peaks; then their average  $\bar{\pi}_B^{(i-1)}$  becomes flat [Fig. 4(h)]. The iteration of such an update leads to a generalist strategy because no concentration occurs and the probability of expressing types is balanced so that the population fitness increases. We will see that the KL term  $\text{KL}^{(i)}$  measures the dissimilarity of the vectors and corresponds to a gain in population fitness by being a generalist strategy. By using the correspondence, we can interpret the vanishing of the KL term when the environment is constant, as a result of non-necessity for being a generalist strategy.

We rewrite Eq. (33) to show that the variance and KL terms,  $\tilde{\Sigma}^{(i)}$  and  $\text{KL}^{(i)}$ , measure the similarity and dissimilarity of the environmental states, respectively. We first see that the variance term measures the similarity between environmental states. The variance term is equal to the following:

$$\tilde{\Sigma}^{(i)} = \langle \log\text{-Cov}_{\pi_F^{(i-1)}} [F_y(x), F_{y'}(x)] \rangle_{Q(y)Q(y')}. \quad (43)$$

Since the log covariance measures the similarity between two environmental states, the variance term measures that between all environmental states. We can say the opposite for the KL term, as it is equal to the following:

$$\text{KL}^{(i)} = \left\langle -\log \frac{\langle F_y(x)F_{y'}(x) \rangle_{\pi_F^{(i-1)}}}{\langle F_y(x) \rangle_{\pi_F^{(i-1)}} \langle F_{y'}(x) \rangle_{\pi_F^{(i-1)}}} \right\rangle_{Q(y)Q(y')}. \quad (44)$$

See Appendix C for the proof. In principle, the KL term is larger when the environmental states are more dissimilar because the second moment  $\langle F_y(x)F_{y'}(x) \rangle_{\pi_F^{(i-1)}}$  appears in the numerator, as long as  $\langle F_y(x) \rangle_{\pi_F^{(i-1)}}$  in the denominator does not change a lot. Therefore, the KL term measures the dissimilarity of the environmental states.

### VIII. NUMERICAL VALIDATION OF FISHER'S FUNDAMENTAL THEOREM FOR ANCESTRAL LEARNING

We numerically verified the FF-thm for ancestral learning. We simulated four different models whose stochastic properties  $Q(y)$  of the environments are different. In each model, we investigated whether the FF-thm holds, that is,  $\Delta_{\text{ac}}\lambda^{(i)} \approx \Delta_{\text{ex}}\lambda^{(i)}$ . The learning rate  $\alpha = 1.0$  unless otherwise specified.

In addition, we set  $\tau_{\text{est}} = 1000$  to suppress the fluctuation of  $j_{\text{emp}}$  [cf. Eq. (51)].

We first validated FF-thm when the environment is constant. We simulated the model shown in Fig. 4(a) and call it a constant environment model. We observed that  $\Delta_{\text{ac}}\lambda^{(i)} \approx \Delta_{\text{ex}}\lambda^{(i)}$  along the lineage of an agent whose initial strategy was  $\pi_F^{(0)} = (0.5, 0.5)$  [Fig. 4(b)]. To check the validity of the FF-thm beyond one lineage, we compared  $\Delta_{\text{ac}}\lambda^{(1)}$  and  $\Delta_{\text{ex}}\lambda^{(1)}$  of the agents, each of which had an initial strategy generated uniformly at random [Fig. 4(c)]. We observed that  $\Delta_{\text{ac}}\lambda^{(1)} \approx \Delta_{\text{ex}}\lambda^{(1)}$  for most of the initial strategies.

We then verified the FF-thm when the environment is not constant by using three simulation models. First, we used the model shown in Fig. 4(d), which has similar environmental states and is thus called a similar environment model. In this model, the optimal strategy is specialized to  $\{0\}$  [Fig. 4(e)], and the variance term  $\tilde{\Sigma}^{(i)}$  is expected to dominate. Figure 4(f) showed  $\Delta_{\text{ac}}\lambda^{(i)}$ ,  $\Delta_{\text{ex}}\lambda^{(i)}$ ,  $\tilde{\Sigma}^{(i)}$ , and  $\text{KL}^{(i)}$  along the lineage of an agent whose initial strategy was  $\pi_F^{(0)} = (0.5, 0.5)$ . From the plot, we found that  $\Delta_{\text{ac}}\lambda^{(i)} \approx \Delta_{\text{ex}}\lambda^{(i)}$  and that the variance term dominated, as expected.

Next, we simulated the model shown in Fig. 4(g). As the environmental states are dissimilar in this model, it is referred to as a dissimilar environment model. In this model, the optimal strategy is a generalist one as illustrated in Fig. 4(h), and the KL term [Eq. (39)] is expected to be non-negligible. Figure 4(i) showed  $\Delta_{\text{ac}}\lambda^{(i)}$ ,  $\Delta_{\text{ex}}\lambda^{(i)}$ ,  $\tilde{\Sigma}^{(i)}$ , and  $\text{KL}^{(i)}$  along the lineage of an agent whose initial strategy was  $\pi_F^{(0)} = (0.9, 0.1)$ . We verified  $\Delta_{\text{ac}}\lambda^{(i)} \approx \Delta_{\text{ex}}\lambda^{(i)}$  and found that the KL term was not negligible, as expected. We also observed that  $|\tilde{\Sigma}^{(i)}| \approx |\text{KL}^{(i)}|$  as  $i$  increases.

Finally, we consider the model shown in Fig. 4(j), where the environmental states 0 and 1 are similar whereas state 2 is dissimilar from them. This model is consequently called a combined model. In this model, the optimal strategy  $\pi^* = (0.72, 0, 0.28)$  is a combination of specialist to  $\mathcal{X}' = \{0, 2\}$  and of generalist in  $\mathcal{X}'$ . Figure 4(k) shows  $\Delta_{\text{ac}}\lambda^{(i)}$ ,  $\Delta_{\text{ex}}\lambda^{(i)}$ ,  $\tilde{\Sigma}^{(i)}$ , and  $\text{KL}^{(i)}$  along the lineage of an agent whose initial strategy was  $(0.05, 0.15, 0.8)$ . We could see that  $\Delta_{\text{ac}}\lambda^{(i)} \approx \Delta_{\text{ex}}\lambda^{(i)}$ . We also observed that the KL term is not negligible. As the variance term dropped faster than the KL term, an agent acquired the specialist strategy first and then acquired the generalist strategy. This interpretation was also supported by the strategy  $\pi_F^{(5)} = (0.31, 0.04, 0.65)$  just before the fifth update, when the variance term became negative for the first time. The strategy was almost specialized to  $\mathcal{X}' = \{0, 2\}$ . In contrast, the strategy was not of generalist in  $\mathcal{X}'$  because  $\pi_F^{(5)}(0)$  and  $\pi_F^{(5)}(2)$  were far from the optimal probabilities  $\pi_F^*(0)$  and  $\pi_F^*(2)$ , respectively. To check the validity of the FF-thm beyond one lineage, we compared  $\Delta_{\text{ac}}\lambda^{(1)}$  and  $\Delta_{\text{ex}}\lambda^{(1)}$  of the agents, each of which had an initial strategy generated uniformly at random [Fig. 4(l)]. We observed that  $\Delta_{\text{ac}}\lambda^{(1)} \approx \Delta_{\text{ex}}\lambda^{(1)}$  for most of the initial strategies.

### IX. TRADE-OFF BETWEEN THE LEARNING RATE AND UPDATE INTERVAL

The FF-thm for ancestral learning is derived for  $\alpha = 1$  and  $\tau_{\text{est}} \gg 1$ . To address other situations, especially one where  $\tau_{\text{est}}$  is not very large, we further extend the FF-thm for ancestral

learning to the cases where  $\alpha < 1.0$  and show that there is a trade-off relationship between  $\alpha$  and  $\tau_{\text{est}}$ . First, we define an  $\log$  covariance by generalizing Eq. (30) as follows:

$$\log\text{-Cov}_p^\alpha [f(x), g(x)] := \log \left( 1 + \alpha \frac{\text{Cov}_p [f(x), g(x)]}{\langle f(x) \rangle_p \langle g(x) \rangle_p} \right). \quad (45)$$

By using this quantity, we have

$$\Delta\lambda^{(i)} = \langle \log\text{-Cov}_{\pi_{\text{F}}^{(i-1)}}^\alpha [e^{k(x,y)}, e^{k(x,y')}] \rangle_{\mathcal{Q}(y)\mathcal{Q}(y')} + \mathcal{D}[\mathcal{Q}(y)\mathcal{Q}(y') \| \bar{\mathcal{Q}}_\alpha^{(i)}(y' | y)\mathcal{Q}(y)], \quad (46)$$

where

$$\bar{\mathcal{Q}}_\alpha^{(i)}(y' | y) \propto \sum_{x \in \mathcal{X}} e^{k(x,y)} \pi_\alpha^{(i-1)}(x | y') \mathcal{Q}(y') \quad (47)$$

and

$$\pi_\alpha^{(i-1)}(x | y') = \alpha \pi_{\text{B}}^{(i-1)}(x | y') + (1 - \alpha) \pi_{\text{F}}^{(i-1)}(x). \quad (48)$$

See Appendix E for the proof. We define the actual and expected gains, which generalize Eqs. (36) and (37), by the left- and right-hand sides of Eq. (46), respectively, as

$$\Delta_{\text{ac}}\lambda^{(i)} := \lambda(\pi_{\text{F}}^{(i)}) - \lambda(\pi_{\text{F}}^{(i-1)}) \quad (49)$$

and

$$\Delta_{\text{ex}}\lambda^{(i)} := \langle \log\text{-Cov}_{\pi_{\text{F}}^{(i-1)}}^\alpha [e^{k(x,y)}, e^{k(x,y')}] \rangle_{\mathcal{Q}(y)\mathcal{Q}(y')} + \mathcal{D}[\mathcal{Q}(y)\mathcal{Q}(y') \| \bar{\mathcal{Q}}_\alpha^{(i)}(y' | y)\mathcal{Q}(y)]. \quad (50)$$

To verify the validity of the FF-thm [Eq. (50)] for  $\alpha < 1.0$ , we simulated the constant environment model [Fig. 4(a)], and the combined model [Fig. 4(j)] when the learning rate is  $\alpha = 0.1$ . We compared  $\Delta_{\text{ac}}\lambda^{(1)}$  and  $\Delta_{\text{ex}}\lambda^{(1)}$  of the agents, each of which had an initial strategy generated uniformly at random [Figs. 4(c) and 4(l)]. We observed  $\Delta_{\text{ac}}\lambda^{(1)} \approx \Delta_{\text{ex}}\lambda^{(1)}$  for most of the strategies.

When  $\tau_{\text{est}} < \infty$ , FF-thm [Eq. (46)] does not hold and  $\Delta_{\text{ac}}\lambda^{(i)} < \Delta_{\text{ex}}\lambda^{(i)}$ . Owing to the finite update interval, the ancestral information  $j_{\text{emp}}^{\pi_{\text{F}}^{(i-1)}}$  and the updated strategy  $\pi_{\text{F}}^{(i)} = \alpha j_{\text{emp}}^{\pi_{\text{F}}^{(i-1)}} + (1 - \alpha) \pi_{\text{F}}^{(i-1)}$  fluctuate around their expectations  $\bar{\pi}_{\text{B}}^{(i-1)}$  and  $\bar{\pi}_\alpha^{(i-1)} = \alpha \bar{\pi}_{\text{B}}^{(i-1)} + (1 - \alpha) \pi_{\text{F}}^{(i-1)}$ , respectively. The average population fitness  $\langle \lambda(\pi_{\text{F}}^{(i)}) \rangle$  with respect to this fluctuation is smaller than  $\lambda(\bar{\pi}_\alpha^{(i-1)})$  owing to the concavity of  $\lambda$  and Jensen's inequality. When  $\tau_{\text{est}}$  is sufficiently large (but still finite), we can quantify this decrease as

$$\Delta_{\text{ac}}\lambda^{(i)} \approx \Delta_{\text{ex}}\lambda^{(i)} + \frac{\alpha^2}{2} \text{Tr}(\mathbf{I}_\lambda \mathbf{V}). \quad (51)$$

Here,  $\text{Tr}(\mathbf{A})$  is the trace of matrix  $\mathbf{A}$ , and the matrix  $\mathbf{V}$  is the covariance matrix of  $j_{\text{emp}}$  defined by

$$\mathbf{V}(x, x') = \langle j_{\text{emp}}(x) j_{\text{emp}}(x') \rangle - \bar{\pi}_{\text{B}}(x) \bar{\pi}_{\text{B}}(x'). \quad (52)$$

In addition,

$$\mathbf{I}_\lambda(x, x') = \frac{\partial^2 \lambda(\bar{\pi}_\alpha)}{\partial \pi(x) \partial \pi(x')}. \quad (53)$$

See Appendix F for the proof. We note that the second term of Eq. (51) is nonpositive because of the negative semidefiniteness of  $\mathbf{I}_\lambda$  shown from the concavity of  $\lambda$ . As  $\mathbf{V}$  is of the

order  $1/\tau_{\text{est}}$ , the deviation  $\alpha^2 \text{Tr}(\mathbf{I}_\lambda \mathbf{V})/2$  from the FF-thm for  $\tau_{\text{est}} = \infty$  is negligible if the learning rate is sufficiently small when compared with the update interval  $\tau_{\text{est}}$ :  $\alpha^2/\tau_{\text{est}} \ll 1$ . Thus, there is a trade-off between  $\alpha$  and  $\tau_{\text{est}}$  in relation to the learning efficiency.

In Sec. VI, we mainly focus on the case of  $\tau_{\text{est}} = \infty$  to make FF-thm [Eq. (46)] intuitive. However, a short  $\tau_{\text{est}}$  is realistic and may be beneficial in both biological and engineering systems. The benefit of a short  $\tau_{\text{est}}$  is that an agent has more opportunities for acceleration by updates of strategy. The drawback is that the acceleration by each update may not always be guaranteed due to the fluctuation of  $j_{\text{emp}}$  around its expectation  $\pi_{\text{B}}$ . As a result, the expected acceleration is decreased. Equation (51) indicates that the decrease is of the order  $\alpha^2/\tau_{\text{est}}$ . This implies that an agent can keep the decrease small by adopting a small  $\alpha$  compared to  $\tau_{\text{est}}$ , although such a small learning rate slows the acceleration by each update [Eq. (46)]. In other words, the decrease in memory size  $\tau_{\text{est}}$  can be compensated by the decrease in learning speed  $\alpha$ . Because of the decrease in acceleration, Eq. (51) depends on the second power of  $\alpha$ , whereas on the first power of  $\tau_{\text{est}}$ , an agent might prefer a pair of small  $\alpha$  and short  $\tau_{\text{est}}$  to that of large  $\alpha$  and long  $\tau_{\text{est}}$ . Indeed, we have numerically shown that ancestral learning accelerated the evolutionary process with a small  $\alpha = 0.01$  and short  $\tau_{\text{est}} = 1$  in Sec. III. In such a situation, our extended FF-thm is insightful because the deviation [second term on the right-hand side of Eq. (51)] is small.

## X. DISCUSSION AND CONCLUSIONS

First, we numerically showed that ancestral learning could accelerate evolutionary processes. We then proved that an agent could estimate the gradient of the population fitness from its ancestral information  $j_{\text{emp}}$ , even without communication between agents. The acceleration was then quantified by extending the FF-thm for ancestral learning and revealed that the gain of population fitness by ancestral learning was associated with the log variance of the individual fitness in a strategy. Finally, we determined the trade-off relationship between the learning rate and update interval. Overall, we have established a theoretical framework with which we can characterize and evaluate the impact of learning on evolutionary processes.

Due to its generality, this framework may be applied to various problems. A biological example is the adaptation (microevolution) of organisms. In the last decade, cell phenotypic variability has been intensely analyzed using experimental techniques to measure the dynamics of the phenotype at the single-cell level, such as division time and cell size [14,15], and sensitivity to chemical substances in chemotaxis [16]. These experimental works have revealed that phenotypic states can be inherited over generations with certain regularities. The results of this study will help to elucidate how the inheritance of phenotypes occurs in cells as a form of learning. Furthermore, our framework may also be employed to develop new methods for measuring and quantifying the impacts of learning from single-cell tracking data.

In engineering, the quantification of how much learning improves the performance of algorithms, such as the memetic algorithms discussed in the Introduction, is a current research

interest. For example, it has recently been utilized in the field of artificial intelligence to integrate reinforcement learning with evolutionary algorithms [39]. While some papers have suggested numerically that the optimization process using the memetic algorithm is faster than with the vanilla genetic algorithm, they lack theoretical quantification and justification of acceleration. Our modified FF-thm can address this problem. With our theory, we may design more efficient algorithms and tune their parameters effectively.

The theory, however, could be improved further, as there are some additional factors that might be useful for agents to learn. One such factor is the type of parent. While an agent with ancestral learning uses the ancestor's type  $j_{\text{emp}}$ , it does not use the type of the parent directly. Such strong dependence on the parent might be beneficial when the environmental state is strongly correlated with the previous state. When the type  $x$  of an agent depends on  $x'$  of the parent, the type-switching strategy should be modeled as a Markov transition  $\mathbb{T}_F(x | x')$  instead of the distribution of  $\pi_F(x)$ . Promising techniques currently available for generalization are large deviations and variational representations, which played an important role in the present study, and Markov chains in random environments [24,25].

Another factor to consider is the communication between agents. Although we showed that the agent can estimate the gradient without communication, learning with such information might further accelerate the evolutionary process when compared to ancestral learning. The acceleration by ancestral learning becomes small when the update interval  $\tau_{\text{est}}$  is short owing to the fluctuations of  $j_{\text{emp}}$  [Eq. (51)]. Communications between agents may thus help to suppress such fluctuations.

The final factor is the sensing of the environmental state. In the context of population dynamics, researchers have considered the situation in which an agent receives a sensing signal  $z$  of the environmental state  $y$  and then expresses their type using a signal-dependent strategy  $\pi_F(x | z)$  [3–6]. Since sensing is another form of information processing, we should consider the unification of sensing and learning to understand the significance of information processing in organisms. In such a setting, an agent might attain the optimal strategy  $\pi_F^*(x | z)$  via extended ancestral learning. To achieve such unification, we need a theory that integrates the prospective and retrospective information obtained by both sensing and learning.

The source code for simulations is available at Ref. [40]. The language was C++17 with Boost 107100. We used Windows Subsystem for Linux 2. The operating system was Ubuntu 20.04.1 LTS on Windows 10 version 2004. We used gcc 9.3.0 for compiling. For Fig. 3, we used graphviz 2.43.0 (0) and colormap [41]. For the other plots, we used matplotlib-cpp [42], which requires PYTHON 3. We used PYTHON 3.8.5.

#### ACKNOWLEDGMENTS

The first author received JSPS Research Fellowship Grant No. JP19J22607 and JST ACT-X Grant No. JPMJAX190L. This research was supported by JSPS KAKENHI Grants No. 19H05799 and No. 19H03216 and by JST CREST Grants No. JPMJCR1927 and No. JPMJCR2011.

#### APPENDIX A: VARIATIONAL REPRESENTATION OF THE POPULATION FITNESS [EQ. (16)]

The proof is a special case of Refs. [6,31]. For the completeness of the paper, we give the proof. For a fixed  $y \in \mathcal{Y}$  and an arbitrary distribution  $\pi$  over  $\mathcal{X}$ ,

$$\log \langle e^{k(x,y)} \rangle_{\pi_F(x)} = \log \sum_{x \in \mathcal{X}} \pi(x) \frac{\pi_F(x)}{\pi(x)} e^{k(x,y)}. \quad (\text{A1})$$

By applying Jensen's inequality, we have

$$\begin{aligned} \log \langle e^{k(x,y)} \rangle_{\pi_F(x)} &\geq \sum_{x \in \mathcal{X}} \pi(x) \left[ \log \frac{\pi_F(x)}{\pi(x)} e^{k(x,y)} \right] \\ &= \sum_{x \in \mathcal{X}} \pi(x) \left[ k(x,y) - \log \frac{\pi(x)}{\pi_F(x)} \right] \\ &= \sum_{x \in \mathcal{X}} \pi(x) k(x,y) - \mathcal{D}[\pi \| \pi_F]. \end{aligned} \quad (\text{A2})$$

By substituting  $\pi(x)$  with  $\pi_B(x | y)$ , we can see that the equality is attained. Therefore,

$$\log \langle e^{k(x,y)} \rangle_{\pi_F(x)} = \max_{\pi} \left\{ \sum_{x \in \mathcal{X}} k(x,y) \pi(x) - \mathcal{D}[\pi \| \pi_F] \right\}. \quad (\text{A3})$$

By averaging the equality with respect to  $Q(y)$ , we have Eq. (16).

#### APPENDIX B: GRADIENT OF THE POPULATION FITNESS [EQS. (18) AND (20)]

The proof is essentially the same as in Ref. [31]. Since the maximizer of the right-hand side of Eq. (A3) is  $\pi_B(x | y)$ ,

$$\log \langle e^{k(x,y)} \rangle_{\pi_F(x)} = \sum_{x \in \mathcal{X}} k(x,y) \pi_B(x | y) - \mathcal{D}[\pi_B \| \pi_F]. \quad (\text{B1})$$

We differentiate both sides with respect to  $\pi_F(x)$  while taking into account the dependence of  $\pi_B$  on  $\pi_F$ :

$$\begin{aligned} \frac{\partial}{\partial \pi_F(x)} \log \langle e^{k(x,y)} \rangle_{\pi_F(x)} &= -\frac{\partial \mathcal{D}[\pi_B \| \pi_F]}{\partial \pi_F(x)} + \sum_{x' \in \mathcal{X}} \frac{\partial \pi_B(x' | y)}{\partial \pi_F(x)} \frac{\partial F[\pi_B]}{\partial \pi_B(x' | y)}, \end{aligned} \quad (\text{B2})$$

where  $F[\pi] := \sum_{x \in \mathcal{X}} k(x,y) \pi(x) - \mathcal{D}[\pi \| \pi_F]$ . Since  $\pi_B$  is the maximizer of  $F$ , the derivative of  $F$  at  $\pi_B$  is zero and consequently the second term vanishes. Therefore,

$$\frac{\partial}{\partial \pi_F(x)} \log \langle e^{k(x,y)} \rangle_{\pi_F(x)} = \frac{\pi_B(x | y)}{\pi_F(x)}. \quad (\text{B3})$$

By taking the average with respect to  $Q(y)$ , we have Eq. (18).

We next prove Eq. (20) via the Lagrange multiplier method. For sufficiently small  $\epsilon$ , it suffices to solve the following linearized optimization:

$$\max_{\delta \pi} \sum_{x \in \mathcal{X}} \frac{\bar{\pi}_B(x)}{\pi_F(x)} \delta \pi(x), \quad (\text{B4})$$

under the constraints  $\sum_{x \in \mathcal{X}} \delta \pi(x) = 0$  and  $\mathcal{D}[\pi_F \| \pi_F + \delta \pi] = \epsilon$ . For a sufficiently small  $\epsilon$ , we can

approximate  $\mathcal{D}[\pi_F \|\pi_F + \delta\pi]$  by using the *Fisher information matrix* [43] as

$$\begin{aligned} \mathcal{D}[\pi_F \|\pi_F + \delta\pi] &\approx \frac{1}{2} \sum_{x, x' \in \mathcal{X}} \delta\pi(x) \delta_{x, x'} \frac{1}{\pi_F(x)} \delta\pi(x') \\ &= \frac{1}{2} \sum_{x \in \mathcal{X}} \frac{\delta\pi^2(x)}{\pi_F(x)}. \end{aligned} \quad (\text{B5})$$

Here, the Fisher information matrix is a  $|\mathcal{X}| \times |\mathcal{X}|$  diagonal matrix with diagonal entries  $\{1/\pi_F(x)\}_{x \in \mathcal{X}}$ . By using this approximation, the Lagrangian function is

$$\begin{aligned} L(\delta\pi; \lambda, \lambda') &= \sum_{x \in \mathcal{X}} \frac{\bar{\pi}_B(x)}{\pi_F(x)} \delta\pi(x) + \lambda \left( \frac{1}{2} \sum_{x \in \mathcal{X}} \frac{\delta\pi^2(x)}{\pi_F(x)} - \epsilon \right) \\ &\quad + \lambda' \left( \sum_{x \in \mathcal{X}} \delta\pi(x) \right). \end{aligned} \quad (\text{B6})$$

By differentiating  $L$  with respect to  $\delta\pi(x)$ , we have the stationary condition:

$$\frac{\partial L}{\partial \delta\pi(x)} = \frac{\bar{\pi}_B(x)}{\pi_F(x)} + \frac{\lambda \delta\pi(x)}{\pi_F(x)} + \lambda' = 0, \quad (\text{B7})$$

for all  $x \in \mathcal{X}$ . By multiplying  $\pi_F(x)$  and taking the sum  $\sum_{x \in \mathcal{X}}$  of both sides of Eq. (B7), we have

$$1 + \lambda' = 0. \quad (\text{B8})$$

We here used  $\sum_{x \in \mathcal{X}} \delta\pi(x) = 0$ . By rearranging Eq. (B7) and substituting  $\lambda' = -1$ , we have

$$\delta\pi(x) = \frac{\pi_F(x) - \bar{\pi}_B(x)}{\lambda} \propto \bar{\pi}_B(x) - \pi_F(x). \quad (\text{B9})$$

### APPENDIX C: FISHER'S FUNDAMENTAL THEOREM OF NATURAL SELECTION [EQS. (26) AND (31)]

We first prove Eq. (26) for the completeness of the paper. By direct calculation,

$$\begin{aligned} \Delta \langle e^{k(x)} \rangle_{p^{(t)}} &= \sum_{x \in \mathcal{X}} e^{k(x)} p^{(t)}(x) - \sum_{x \in \mathcal{X}} e^{k(x)} p^{(t-1)}(x) \\ &= \sum_{x \in \mathcal{X}} e^{k(x)} \frac{e^{k(x)} p^{(t-1)}(x)}{\sum_{x' \in \mathcal{X}} e^{k(x')} p^{(t-1)}(x')} \\ &\quad - \sum_{x \in \mathcal{X}} e^{k(x)} p^{(t-1)}(x) \\ &= \frac{\sum_{x \in \mathcal{X}} (e^{k(x)})^2 p^{(t-1)}(x) - (\sum_{x \in \mathcal{X}} e^{k(x)} p^{(t-1)}(x))^2}{\sum_{x \in \mathcal{X}} e^{k(x)} p^{(t-1)}(x)} \\ &= \frac{\mathbb{V}_{p^{(t-1)}}[e^{k(x)}]}{\langle e^{k(x)} \rangle_{p^{(t-1)}}}. \end{aligned} \quad (\text{C1})$$

We next prove Eq. (31). By direct calculation,

$$\begin{aligned} \Delta \lambda^{(t)} &= \log \sum_{x \in \mathcal{X}} e^{k(x)} p^{(t)}(x) - \log \sum_{x \in \mathcal{X}} e^{k(x)} p^{(t-1)}(x) \\ &= \log \sum_{x \in \mathcal{X}} e^{k(x)} \frac{e^{k(x)} p^{(t-1)}(x)}{\sum_{x' \in \mathcal{X}} e^{k(x')} p^{(t-1)}(x')} \end{aligned}$$

$$\begin{aligned} &- \log \sum_{x \in \mathcal{X}} e^{k(x)} p^{(t-1)}(x) \\ &= \log \sum_{x \in \mathcal{X}} (e^{k(x)})^2 p^{(t-1)}(x) - 2 \log \sum_{x \in \mathcal{X}} e^{k(x)} p^{(t-1)}(x) \\ &= \log \frac{\langle (e^{k(x)})^2 \rangle_{p^{(t-1)}}}{\langle e^{k(x)} \rangle_{p^{(t-1)}}^2} \\ &= \log -\mathbb{V}_{p^{(t-1)}}[e^{k(x)}]. \end{aligned} \quad (\text{C2})$$

### APPENDIX D: FISHER'S FUNDAMENTAL THEOREM OF ANCESTRAL LEARNING FOR NONCONSTANT ENVIRONMENT [EQS. (33) AND (44)]

We first prove Eq. (33). By direct calculation,

$$\begin{aligned} \lambda(\pi_F^{(i)}) &= \langle \log \langle e^{k(x, y)} \rangle_{\bar{\pi}_B^{(i-1)}} \rangle_{Q(y)} \\ &= \langle \log \langle e^{k(x, y)} \rangle_{\bar{\pi}_B^{(i-1)}} \rangle_{Q(y)Q(y')} \\ &= \langle \log \langle e^{k(x, y)} \rangle_{\pi_B^{(i-1)}(x|y')} \rangle_{Q(y)Q(y')} \\ &\quad + \left\langle \log \frac{\langle e^{k(x, y)} \rangle_{\bar{\pi}_B^{(i-1)}}}{\langle e^{k(x, y)} \rangle_{\pi_B^{(i-1)}(x|y')}} \right\rangle_{Q(y)Q(y')}. \end{aligned} \quad (\text{D1})$$

We first treat the first term. By a similar argument to Eq. (31), the term inside the expectation satisfies the following relationship:

$$\begin{aligned} &\log \langle e^{k(x, y)} \rangle_{\pi_B^{(i-1)}(x|y')} - \log \langle e^{k(x, y)} \rangle_{\pi_F^{(i-1)}} \\ &= \log \sum_{x \in \mathcal{X}} e^{k(x, y)} \frac{e^{k(x, y')} \pi_F^{(i-1)}(x)}{\langle e^{k(x', y')} \rangle_{\pi_F^{(i-1)}(x')}} \\ &\quad - \log \langle e^{k(x, y)} \rangle_{\pi_F^{(i-1)}} \\ &= \log \frac{\langle e^{k(x, y) + k(x, y')} \rangle_{\pi_F^{(i-1)}}}{\langle e^{k(x, y)} \rangle_{\pi_F^{(i-1)}} \langle e^{k(x, y')} \rangle_{\pi_F^{(i-1)}}} \\ &= \log -\text{Cov}_{\pi_F^{(i-1)}}[e^{k(x, y)}, e^{k(x, y')}] \end{aligned} \quad (\text{D2})$$

By taking the average with respect to  $Q(y)Q(y')$ , we have

$$\begin{aligned} &\langle \log \langle e^{k(x, y)} \rangle_{\pi_B^{(i-1)}(x|y')} \rangle_{Q(y)Q(y')} - \lambda(\pi_F^{(i-1)}) \\ &= \langle \log -\text{Cov}_{\pi_F^{(i-1)}}[e^{k(x, y)}, e^{k(x, y')}] \rangle_{Q(y)Q(y')}. \end{aligned} \quad (\text{D3})$$

We next treat the second term of Eq. (D1). By definition,

$$\begin{aligned} \frac{\bar{Q}^{(i)}(y' | y)}{Q(y')} &= \frac{\sum_{x \in \mathcal{X}} e^{k(x, y)} \pi_B^{(i-1)}(x | y')}{\sum_{x \in \mathcal{X}, y' \in \mathcal{Y}} e^{k(x, y)} \pi_B^{(i-1)}(x | y') Q(y')} \\ &= \frac{\langle e^{k(x, y)} \rangle_{\pi_B^{(i-1)}(x|y')}}{\langle e^{k(x, y)} \rangle_{\bar{\pi}_B^{(i-1)}}}. \end{aligned} \quad (\text{D4})$$

We thus have

$$\begin{aligned} &\left\langle \log \frac{\langle e^{k(x, y)} \rangle_{\bar{\pi}_B^{(i-1)}}}{\langle e^{k(x, y)} \rangle_{\pi_B^{(i-1)}(x|y')}} \right\rangle_{Q(y)Q(y')} \\ &= \left\langle \log \frac{Q(y')}{\bar{Q}^{(i)}(y' | y)} \right\rangle_{Q(y)Q(y')} \end{aligned}$$

$$\begin{aligned}
&= \left\langle \log \frac{Q(y)Q(y')}{\bar{Q}^{(i)}(y' | y)Q(y)} \right\rangle_{Q(y)Q(y')} \\
&= \mathcal{D}[Q(y)Q(y') \| \bar{Q}^{(i)}(y' | y)Q(y)]. \quad (\text{D5})
\end{aligned}$$

In conclusion, we proved Eq. (33).

We next prove Eq. (44). By Eq. (D4),

$$\begin{aligned}
\log \frac{Q(y)Q(y')}{\bar{Q}^{(i)}(y' | y)Q(y)} &= -\log \frac{\langle e^{k(x,y)} \rangle_{\pi_B^{(i-1)}(x|y')}}{\langle e^{k(x,y)} \rangle_{\bar{\pi}_B^{(i-1)}}} \\
&= -\log \frac{\langle e^{k(x,y)+k(x,y')} \rangle_{\pi_F^{(i-1)}}}{\langle e^{k(x,y)} \rangle_{\bar{\pi}_B^{(i-1)}} \langle e^{k(x,y')} \rangle_{\pi_F^{(i-1)}}} \\
&= -\log \frac{\langle F_y(x)F_{y'}(x) \rangle_{\pi_F^{(i-1)}}}{\langle F_y(x) \rangle_{\bar{\pi}_B^{(i-1)}} \langle F_{y'}(x) \rangle_{\pi_F^{(i-1)}}}. \quad (\text{D6})
\end{aligned}$$

By averaging with respect to  $Q(y)Q(y')$ , we have Eq. (44).

#### APPENDIX E: FISHER'S FUNDAMENTAL THEOREM OF ANCESTRAL LEARNING WHEN $\alpha < 1$ [EQ. (46)]

We can prove Eq. (46) by almost the same argument as Eq. (33). Recall that  $\bar{\pi}_\alpha^{(i-1)} = \alpha \bar{\pi}_B^{(i-1)} + (1-\alpha)\pi_F^{(i-1)}$ . By direct calculation, we have

$$\begin{aligned}
\lambda(\pi_F^{(i)}) &= \langle \log \langle e^{k(x,y)} \rangle_{\bar{\pi}_\alpha^{(i-1)}} \rangle_{Q(y)} \\
&= \langle \log \langle e^{k(x,y)} \rangle_{\bar{\pi}_\alpha^{(i-1)}} \rangle_{Q(y)Q(y')} \\
&= \langle \log \langle e^{k(x,y)} \rangle_{\pi_\alpha^{(i-1)}(x|y')} \rangle_{Q(y)Q(y')} \\
&\quad + \left\langle \log \frac{\langle e^{k(x,y)} \rangle_{\bar{\pi}_\alpha^{(i-1)}}}{\langle e^{k(x,y)} \rangle_{\pi_\alpha^{(i-1)}(x|y')}} \right\rangle_{Q(y)Q(y')}. \quad (\text{E1})
\end{aligned}$$

We first treat the first term. By a similar argument to Eq. (31), the term inside the expectation satisfies

$$\begin{aligned}
&\log \langle e^{k(x,y)} \rangle_{\pi_\alpha^{(i-1)}(x|y')} - \log \langle e^{k(x,y)} \rangle_{\pi_F^{(i-1)}} \\
&= \log \left\langle e^{k(x,y)} \left( \alpha \frac{e^{k(x,y')}}{\langle e^{k(x',y')} \rangle_{\pi_F^{(i-1)}(x')}} + 1 - \alpha \right) \right\rangle_{\pi_F^{(i-1)}(x)} \\
&\quad - \log \langle e^{k(x,y)} \rangle_{\pi_F^{(i-1)}} \\
&= \log \left( \alpha \frac{\langle e^{k(x,y)+k(x,y')} \rangle_{\pi_F^{(i-1)}}}{\langle e^{k(x,y)} \rangle_{\pi_F^{(i-1)}}} + (1-\alpha) \langle e^{k(x,y)} \rangle_{\pi_F^{(i-1)}} \right) \\
&\quad - \log \langle e^{k(x,y)} \rangle_{\pi_F^{(i-1)}} \\
&= \log \left( \alpha \frac{\langle e^{k(x,y)+k(x,y')} \rangle_{\pi_F^{(i-1)}}}{\langle e^{k(x,y)} \rangle_{\pi_F^{(i-1)}} \langle e^{k(x,y')} \rangle_{\pi_F^{(i-1)}}} + 1 - \alpha \right). \quad (\text{E2})
\end{aligned}$$

By a similar argument to Eq. (30),

$$\begin{aligned}
&\log \langle e^{k(x,y)} \rangle_{\pi_\alpha^{(i-1)}(x|y')} - \log \langle e^{k(x,y)} \rangle_{\pi_F^{(i-1)}} \\
&= \log \left( \alpha \frac{\text{Cov}_{\pi_F^{(i-1)}}[e^{k(x,y)}, e^{k(x,y')}]}{\langle e^{k(x,y)} \rangle_{\pi_F^{(i-1)}} \langle e^{k(x,y')} \rangle_{\pi_F^{(i-1)}}} + 1 \right) \\
&= \log\text{-Cov}_{\pi_F^{(i-1)}}^\alpha [e^{k(x,y)}, e^{k(x,y')}]. \quad (\text{E3})
\end{aligned}$$

By taking the average with respect to  $Q(y)Q(y')$ , we have

$$\begin{aligned}
&\langle \log \langle e^{k(x,y)} \rangle_{\pi_\alpha^{(i-1)}(x|y')} \rangle_{Q(y)Q(y')} - \lambda(\pi_F^{(i-1)}) \\
&= \langle \log\text{-Cov}_{\pi_F^{(i-1)}}^\alpha [e^{k(x,y)}, e^{k(x,y')}] \rangle_{Q(y)Q(y')}. \quad (\text{E4})
\end{aligned}$$

We next treat the second term of Eq. (E1). By definition,

$$\begin{aligned}
\frac{\bar{Q}_\alpha^{(i)}(y' | y)}{Q(y')} &= \frac{\sum_{x \in \mathcal{X}} e^{k(x,y)} \pi_\alpha^{(i-1)}(x | y')}{\sum_{x \in \mathcal{X}, y' \in \mathcal{Y}} e^{k(x,y)} \pi_\alpha^{(i-1)}(x | y') Q(y')} \\
&= \frac{\langle e^{k(x,y)} \rangle_{\pi_\alpha^{(i-1)}(x|y')}}{\langle e^{k(x,y)} \rangle_{\pi_\alpha^{(i-1)}}}. \quad (\text{E5})
\end{aligned}$$

Thus,

$$\begin{aligned}
&\left\langle \log \frac{\langle e^{k(x,y)} \rangle_{\bar{\pi}_\alpha^{(i-1)}}}{\langle e^{k(x,y)} \rangle_{\pi_\alpha^{(i-1)}(x|y')}} \right\rangle_{Q(y)Q(y')} \\
&= \left\langle \log \frac{Q(y')}{\bar{Q}_\alpha^{(i)}(y' | y)} \right\rangle_{Q(y)Q(y')} \\
&= \left\langle \log \frac{Q(y)Q(y')}{\bar{Q}_\alpha^{(i)}(y' | y)Q(y)} \right\rangle_{Q(y)Q(y')} \\
&= \mathcal{D}[Q(y)Q(y') \| \bar{Q}_\alpha^{(i)}(y' | y)Q(y)]. \quad (\text{E6})
\end{aligned}$$

In conclusion, we proved Eq. (46).

#### APPENDIX F: FISHER'S FUNDAMENTAL THEOREM OF ANCESTRAL LEARNING WHEN $\tau_{\text{est}}$ IS FINITE [EQ. (51)]

When  $\tau_{\text{est}}$  is sufficiently large (but finite), we can approximate  $j_{\text{emp}}$  by the central limit theorem [44] as

$$j_{\text{emp}} \sim \mathcal{N}(\bar{\pi}_B^{(i-1)}, \mathbf{V}), \quad (\text{F1})$$

where  $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  is the multivariate normal distribution with mean  $\boldsymbol{\mu}$  and covariance  $\boldsymbol{\Sigma}$ . The updated strategy  $\pi_F^{(i)} = \alpha j_{\text{emp}} + (1-\alpha)\pi_F^{(i-1)}$  satisfies

$$\pi_F^{(i)} \sim \mathcal{N}(\bar{\pi}_\alpha, \alpha^2 \mathbf{V}), \quad (\text{F2})$$

where we omit the superscript of  $\bar{\pi}_\alpha^{(i-1)}$  to avoid the complication. The growth rate is approximated as

$$\begin{aligned}
\lambda(\pi_F^{(i)}) &= \lambda(\bar{\pi}_\alpha + \delta\pi) \\
&\approx \lambda(\bar{\pi}_\alpha) + \sum_{x \in \mathcal{X}} \frac{\partial \lambda(\bar{\pi}_\alpha)}{\partial \pi(x)} \delta\pi(x) \\
&\quad + \frac{1}{2} \sum_{x, x' \in \mathcal{X}} \delta\pi(x) \frac{\partial^2 \lambda(\bar{\pi}_\alpha)}{\partial \pi(x) \partial \pi(x')} \delta\pi(x') \\
&= \lambda(\bar{\pi}_\alpha) + \sum_{x \in \mathcal{X}} \frac{\partial \lambda(\bar{\pi}_\alpha)}{\partial \pi(x)} \delta\pi(x) \\
&\quad + \frac{1}{2} \sum_{x, x' \in \mathcal{X}} \delta\pi(x) I_\lambda(x, x') \delta\pi(x'), \quad (\text{F3})
\end{aligned}$$

where  $\delta\pi = \pi_F^{(i)} - \bar{\pi}_\alpha = \alpha j_{\text{emp}}$ . We note that  $\delta\pi \sim \mathcal{N}(\mathbf{0}, \alpha^2 V)$ . By this approximation,

$$\begin{aligned} \Delta_{\text{ac}}\lambda^{(i)} &= \langle \lambda(\pi_F^{(i)}) \rangle_{\mathcal{N}(\bar{\pi}_\alpha, \alpha^2 V)} - \lambda^{(i-1)} \\ &\approx \lambda(\bar{\pi}_\alpha) - \lambda^{(i-1)} + \left\langle \sum_{x \in \mathcal{X}} \frac{\partial \lambda(\bar{\pi}_\alpha)}{\partial \pi(x)} \delta\pi(x) \right\rangle_{\mathcal{N}(\mathbf{0}, \alpha^2 V)} \\ &\quad + \frac{1}{2} \left\langle \sum_{x, x' \in \mathcal{X}} \delta\pi(x) I_\lambda(x, x') \delta\pi(x') \right\rangle_{\mathcal{N}(\mathbf{0}, \alpha^2 V)} \\ &= \Delta_{\text{ex}}\lambda^{(i)} + \frac{1}{2} \left\langle \sum_{x, x' \in \mathcal{X}} \delta\pi(x) I_\lambda(x, x') \delta\pi(x') \right\rangle_{\mathcal{N}(\mathbf{0}, \alpha^2 V)}. \end{aligned} \tag{F4}$$

In the last equation, the third term vanishes because

$$\begin{aligned} &\left\langle \sum_{x \in \mathcal{X}} \frac{\partial \lambda(\bar{\pi}_\alpha)}{\partial \pi(x)} \delta\pi(x) \right\rangle_{\mathcal{N}(\mathbf{0}, \alpha^2 V)} \\ &= \sum_{x \in \mathcal{X}} \frac{\partial \lambda(\bar{\pi}_\alpha)}{\partial \pi(x)} \langle \delta\pi(x) \rangle_{\mathcal{N}(\mathbf{0}, \alpha^2 V)} \\ &= 0. \end{aligned} \tag{F5}$$

By the usual matrix calculation [45],

$$\begin{aligned} &\left\langle \sum_{x, x'} \delta\pi(x) I_\lambda(x, x') \delta\pi(x') \right\rangle_{\mathcal{N}(\mathbf{0}, \alpha^2 V)} \\ &= \alpha^2 \text{Tr}(I_\lambda V). \end{aligned} \tag{F6}$$

In all, we proved Eq. (51).

---

[1] L. A. Urry, M. L. Cain, S. A. Wasserman, and P. V. Minorsky, *Campbell Biology* (Pearson, New York, 2016).

[2] T. Back, *Evolutionary Algorithms in Theory and Practice: Evolution Strategies, Evolutionary Programming, Genetic Algorithms* (Oxford University Press, Oxford, 1996).

[3] P. Haccou and Y. Iwasa, Optimal mixed strategies in stochastic environments, *Theor. Popul. Biol.* **47**, 212 (1995).

[4] E. Kussell and S. Leibler, Phenotypic diversity, population growth, and information in fluctuating environments, *Science* **309**, 2075 (2005).

[5] O. Rivoire and S. Leibler, The value of information for populations in varying environments, *J. Stat. Phys.* **142**, 1124 (2011).

[6] T. J. Kobayashi and Y. Sughiyama, Fluctuation Relations of Fitness and Information in Population Dynamics, *Phys. Rev. Lett.* **115**, 238102 (2015).

[7] B. Xue and S. Leibler, Evolutionary learning of adaptation to varying environments through a transgenerational feedback, *Proc. Natl. Acad. Sci. USA* **113**, 11266 (2016).

[8] T. J. Kobayashi and Y. Sughiyama, Fitness gain of individually sensed information by cells, *Entropy* **21**, 1002 (2019).

[9] R. Dawkins, *The Selfish Gene* (Oxford University Press, Oxford, 2016).

[10] Y. Taniguchi, P. J. Choi, G.-W. Li, H. Chen, M. Babu, J. Hearn, A. Emili, and X. S. Xie, Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells, *Science* **329**, 533 (2010).

[11] A. M. Cunningham, D. M. Walker, A. Ramakrishnan, M. A. Doyle, R. C. Bagot, H. M. Cates, C. J. Peña, O. Issler, C. Lardner, C. Browne, S. J. Russo, L. Shen, and E. J. Nestler, Sperm transcriptional state associated with paternal transmission of stress phenotypes, *J. Neurosci.* **41**, 6202 (2021).

[12] R. Nagpal, H. Tsuji, T. Takahashi, K. Kawashima, S. Nagata, K. Nomoto, and Y. Yamashiro, Sensitive quantitative analysis of the meconium bacterial microbiota in healthy term infants born vaginally or by cesarean section, *Front. Microbiol.* **7**, 1997 (2016).

[13] K. Yoshida, S.-I. Fujita, A. Isotani, T. Kudo, S. Takahashi, M. Ikawa, D. Shiba, M. Shirakawa, M. Muratani, and S. Ishii, Intergenerational effect of short-term spaceflight in mice, *iScience* **24**, 102773 (2021).

[14] P. Wang, L. Robert, J. Pelletier, W. L. Dang, F. Taddei, A. Wright, and S. Jun, Robust growth of Escherichia coli, *Curr. Biol.* **20**, 1099 (2010).

[15] M. Hashimoto, T. Nozoe, H. Nakaoka, R. Okura, S. Akiyoshi, K. Kaneko, E. Kussell, and Y. Wakamoto, Noise-driven growth rate gain in clonal cellular populations, *Proc. Natl. Acad. Sci. USA* **113**, 3251 (2016).

[16] T. L. Min, P. J. Mears, I. Golding, and Y. R. Chemla, Chemotactic adaptation kinetics of individual Escherichia coli cells, *Proc. Natl. Acad. Sci. USA* **109**, 9869 (2012).

[17] P. Moscato *et al.*, On evolution, search, optimization, genetic algorithms and martial arts: Towards memetic algorithms, Caltech Concurrent Computation Program (C3P) Report No. 826 (1989).

[18] Y. Ollivier, L. Arnold, A. Auger, and N. Hansen, Information-geometric optimization algorithms: A unifying picture via invariance principles, *J. Mach. Learn. Res.* **18**, 1 (2017).

[19] J. Seger, What is bet-hedging? in *Oxford Surveys in Evolutionary Biology*, edited by P. H. Harvey and L. Partridge (Oxford University Press, Oxford, UK, 1987), Vol. 4, pp. 182–211.

[20] I. G. de Jong, P. Haccou, and O. P. Kuipers, Bet hedging or not? A guide to proper classification of microbial survival strategies, *BioEssays* **33**, 215 (2011).

[21] If the second or higher derivative is used, then the learning is called the second (or higher) order.

[22] R. A. Fisher, *The Genetical Theory of Natural Selection* (Clarendon Press, Oxford, 1930).

[23] The individual fitness is the mean number of daughters. We use the term “fitness” to follow the terminology of FF-thm [22].

[24] T. Seppalainen, Large deviations for Markov chains with random transitions, *Ann. Probab.* **22**, 713 (1994).

[25] Y. Kifer, Perron-Frobenius theorem, large deviations, and random perturbations in random environments, *Math. Z.* **222**, 677 (1996).

[26] The population fitness is also called the growth rate.

[27] J. A. Hertz, A. Krogh, and R. Palmer, *Introduction to the Theory of Neural Computation*, 1st ed. (CRC Press, Boca Raton, FL, 1991).



- [28] J. Hermisson, O. Redner, H. Wagner, and E. Baake, Mutation-selection balance: Ancestry, load, and maximum principle, *Theor. Popul. Biol.* **62**, 9 (2002).
- [29] E. Baake and H.-O. Georgii, Mutation, selection, and ancestry in branching models: A variational approach, *J. Math. Biol.* **54**, 257 (2007).
- [30] H.-O. Georgii and E. Baake, Supercritical multitype branching processes: The ancestral types of typical individuals, *Adv. Appl. Probab.* **35**, 1090 (2003).
- [31] Y. Sughiyama, T. J. Kobayashi, K. Tsumura, and K. Aihara, Pathwise thermodynamic structure in population dynamics, *Phys. Rev. E* **91**, 032120 (2015).
- [32] T. E. Harris, *The Theory of Branching Process* (RAND Corporation, Santa Monica, CA, 1964).
- [33] E. W. Weissner, Multitype branching processes in random environments, *J. Appl. Probab.* **8**, 17 (1971).
- [34] S. M. Rosenberg, Evolving responsively: Adaptive mutation, *Nat. Rev. Genet.* **2**, 504 (2001).
- [35] T. M. Cover, *Elements of Information Theory* (Wiley, New York, 1999).
- [36] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications* (Springer-Verlag, Berlin, 2010).
- [37] N. Parikh and S. Boyd, Proximal algorithms, *Found. Trends Optim.* **1**, 127 (2014).
- [38] The FF-thm for natural selection is sometimes formulated in a continuous-time model. Here, we present a discrete-time version to see the connection to our ancestral learning.
- [39] S. Khadka, S. Majumdar, T. Nassar, Z. Dwiell, E. Tumer, S. Miret, Y. Liu, and K. Tumer, Collaborative evolutionary reinforcement learning, in *International Conference on Machine Learning* (Proceedings of Machine Learning Research, 2019), pp. 3341–3350.
- [40] [https://github.com/so-nakashima/learning\\_in\\_growing\\_systems](https://github.com/so-nakashima/learning_in_growing_systems)
- [41] <https://github.com/jgreitemann/colormap>
- [42] <https://github.com/lava/matplotlib-cpp>
- [43] S.-I. Amari, *Information Geometry and Its Applications* (Springer, Berlin, 2016), Vol. 194.
- [44] A. W. v. d. Vaart, *Asymptotic Statistics*, Cambridge Series in Statistical and Probabilistic Mathematics (Cambridge University Press, Cambridge, 1998).
- [45] K. B. Petersen and M. S. Pedersen, *The Matrix Cookbook*, version 15 November (Technical University of Denmark, 2012).