PLOS ONE

RESEARCH ARTICLE

How learning can change the course of evolution

Leonel Aguilar[‡*], **Stefano Bennati**[‡], **Dirk Helbing**[¤]

Professorship of Computational Social Science, ETH Zürich, Zürich, Switzerland

¤ Current address: Professorship of Computational Social Science, ETH Zürich, Zürich, Switzerland
‡ S.B. and L.A. are Joint First Authors
* leonel.aguilar@gess.ethz.ch

OPEN ACCESS

Citation: Aguilar L, Bennati S, Helbing D (2019) How learning can change the course of evolution. PLoS ONE 14(9): e0219502. https://doi.org/10.1371/journal.pone.0219502

Editor: Denis Horvath, Pavol Jozef Safarik University in Kosice, SLOVAKIA

Received: December 15, 2018

Accepted: June 25, 2019

Published: September 5, 2019

Data Availability Statement: The source code and data to reproduce this work are available from Github at the following link: https://github.com/leaguilar/baldwin_veering/.

Funding: The authors acknowledge support by the European Commission through the European Research Council Advanced Investigator Grant 'Momentum' (Grant No. 324247). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Abstract

The interaction between phenotypic plasticity, e.g. learning, and evolution is an important topic both in Evolutionary Biology and Machine Learning. The evolution of learning is commonly studied in Evolutionary Biology, while the use of an evolutionary process to improve learning is of interest to the field of Machine Learning. This paper takes a different point of view by studying the effect of learning on the evolutionary process, the so-called Baldwin effect. A well-studied result in the literature about the Baldwin effect is that learning affects the speed of convergence of the evolutionary process towards some genetic configuration, which corresponds to the environment-induced plastic response. This paper demonstrates that learning can change the outcome of evolution, i.e., lead to a genetic configuration that does not correspond to the plastic response. Results are obtained both analytically and experimentally by means of an agent-based model of a foraging task, in an environment where the distribution of resources follows seasonal cycles and the foraging success on different resource types is conditioned by trade-offs that can be evolved and learned. This paper attempts to answer a question that has been overlooked: whether learning has an effect on what genotypic traits are evolved, i.e. the selection of a trait that enables a plastic response changes the selection pressure on a different trait, in what could be described as co-evolution between different traits in the same genome.

1 Introduction

The so called Baldwin effect [1] is a much debated theory in the literature of evolution [2] about how new features are inherited by an individual with phenotypic plasticity [3–5]. Baldwin proposed this new "factor in evolution" [1] to explain how complex features such as an eye can evolve [6–8], as an alternative to the then-popular Lamarckian evolution which assumed that traits acquired by an individual through phenotypic plasticity would be transferred directly to its offspring's genome [9]. This idea went unnoticed until the late 1990s, when it caught the interest of the fields of Psychology, in reference to the evolution of human learning, and Computer Science, in reference to evolutionary computation, machine learning, and

artificial life. Only from the mid-2000s did the Baldwin effect start taking ground in the field of Evolutionary Biology. [10].

Given the long debate surrounding the Baldwin effect, there are different definitions of it with different levels of generality, e.g. "The Baldwin Effect, states that learned behavior and characteristics at the level of individuals can significantly affect evolution at the level of species" [11], Schull relates the Baldwin effect to statements such as "individual developmental responses will necessarily lead to directed and non-random evolutionary change" [12]. The open peer commentaries of [12] highlight different conflicting stances regarding the Baldwin effect and its definition. The working definition of the Baldwin effect used in this paper is: plasticity is a "positive driving force of evolution" that affects the selection pressure such that "standing genetic variation can be selected upon so that evolution can proceed in the direction of the induced plastic response" [8]. According to this definition, the Baldwin effect describes the evolution of a "target" genotypic trait that corresponds to the environment-induced plastic response at the phenotypic level. In other words, the induced plastic response determines the direction towards which the genotype evolves. This definition is especially relevant when considering biologically inspired optimization techniques [13].

A well-known example of the Baldwin effect is that learning, i.e. an instance of phenotypic plasticity, affects the evolutionary process by either speeding up or slowing down the evolution of the "target" genetic configuration.

This work demonstrates that this definition is too restrictive, as a genotypic trait is shown to evolve that differs from the environment-induced plastic response. The term *Baldwin veering effect* is introduced to refer to this new finding and defined as follows: a change in the selection pressure of genetic variations, caused by phenotypic plasticity and induced plastic responses, leads evolution in a different direction from that indicated by the induced plastic response. In other words, the Baldwin veering effect happens when a trait evolves by effect of plasticity that does not correspond to the environment-induced plastic response.

In order to demonstrate the existence of the Baldwin veering effect, the following two conditions have to be verified:

- A trait evolves that differs from the induced plastic response, i.e. the genome and the phenotype converge towards different trait values.
- The evolution of such trait is caused by plasticity, i.e. the genome converges towards different trait values in presence or absence of plasticity.

The effect of plasticity—we choose learning among many potential mechanisms, e.g. polyphenism [14]—on evolution is studied both computationally by means of an agent-based model of a foraging task, modeled after previous work [15, 16], and analytically by means of a mathematical model [17]. Computational experiments and analytical results in a cyclically-changing environment demonstrate the existence of both the Baldwin effect and of the Baldwin veering effect. Specifically, it is found that in a quickly-changing cyclical environment, learning agents evolve a generalist foraging strategy that allows them to adapt quickly to changes in the resource distribution. A generalist configuration is never induced, i.e. learned, at the phenotypic level. Analytical results confirm that plasticity changes the fitness landscape in a way that makes a generalist configuration a global optimum in the space of genotypes.

The novelty of this result is to expand the understanding of the effect of plasticity on evolution by demonstrating that plasticity can affect both the speed and the outcome of evolution. A fundamental difference of this result from previous work [14, 18, 19] is that learning is not only shown to change the phenotype but the genotype as well.

The main contributions of this paper are to show that in a cyclically-changing environment: (I) the well-known Baldwin effect is present, (II) the novel *Baldwin veering effect* is present, (III) a mathematical model captures this new effect and confirms the experimental findings, and (IV) the existence of this new effect depends only on the relation between the speed of learning and the frequency of change in the environment.

2 Methods

The computational model follows the agent-based methodology [20] by studying the interactions of a population of software agents, subject to an evolutionary process [21], that perform a foraging task [15, 22], i.e. search the environment for food in a grid like environment. This model builds on the extensive research in the artificial life community, where software agents have been provided with learning mechanisms [11, 23–26] in an evolutionary context. The time-step driven simulation model is based on previous work [27] and favors simplicity over realism. Modeling realistic entities and ecosystems is outside the scope of this work.

This section provides an overview of the essential components of the computational model, the possible phenomenon occurring at every simulated abstract time-unit, and a detailed description of the environment, agents, decision making, and evaluation metrics. Subsection 2.1 describes the cyclically-changing environment, the resources to be foraged and their seasonality. Subsection 2.2 describes the agent most relevant parameters (aptitude and skill) and the trade-offs that these parameters cause during foraging, the relation between foraging and energy levels, and how energy levels affect fitness, reproduction, and death. Subsection 2.3 details the agent behavior (reactive and learning); additionally, it is explained how reproduction affects the parameters related to the decision making process. Finally, subsection 2.4 defines the measures used to evaluate the agents' behavior. Table 1 contains an overview of the notation used throughout the description of the model. The results presented in this paper are the outcome of 300 Monte Carlo type simulations for each specific scenario.

2.1 Environment

The environment is modeled as a square grid of size $m \times m$ with continuous boundary conditions in which agents can move. Every grid cell can contain one of the two resource types, i.e. $|R| = 2$, whose proportions vary over time [28] such that in every "season" a specific resource is more abundant than the other.

2.1.1 Food sources. The number of cells with resources, $|F^t|$, is constant at every point in time: whenever one cell is emptied, a random quantity $\phi^t_{i,r}$ of resources of the same type spawns at a random location. New food sources are initialized to contain a random quantity of food, driven by the parameter Φ that determines the abundance of food.

2.1.2 Seasons. The environment cycles periodically between two different configurations, named *seasons* [14, 28], which determine what resources are available for agents to forage. Foraging of different resource types is subject to trade-offs: the more an agent specializes in the gathering and consumption of one resource, the less effectively it forages the other resource, e.g. due to neophobia [29], a non-transferable skill set or other constraints, e.g. energy or memory constraints. This trade-off is modeled by a single *skill* parameter that determines the probability of success of foraging two resource types [30]. Environmental change is a known requirement for the evolution of learning, and seasons offer enough predictability for learning to be effective [31].

Table 1. Summary of the mathematical notation used in order of appearance in the text. Notation used for indexes has been slightly abused for the sake of brevity.

Math symbol	Description
$\mathcal{A} = \{a_0, ..., a_N\}$	The set of all N agents ever alive in the simulation
$R = \{r_0, \ldots, r_M\}$	The set of M resource types
$F^t = \{\phi^t_{i,r} \in E^t : \phi^t_{i,r} > 0\}$	Set of all cells containing resources
Φ	The maximum quantity of resource that any cell can contain
$\phi^t_{i,r} \in \mathbb{N}_{\geq 0}$	The quantity of resources of type r in cell i at time t
$E^t = \{\phi^t_{i,r} : 1 \leq i \leq m \times m, r \in R\}$	The configuration of the environment at time t
$T = \{t \in \mathbb{N}_{0,\leq L}\}$	The time steps, t of the simulation
$s^t_a \in [0,1] \;:\; a \in \mathcal{A}^t$	The skill level of agent a at time t
$\mathcal{A}^t = \{a \in \mathcal{A} : a \text{ is alive at timestep } t\}$	The population at time t
$f(a,t) : \mathcal{A}^t \times T \rightarrow \mathbb{R}$	The fitness function
$\epsilon \in \mathbb{R}$	Energy level increased by successful foraging
$g(a, s^t_a, r) : \mathcal{A}^t \times \mathbb{R}_{\geq 0,\leq 1} \times R \rightarrow \{0,1\}$	The foraging success function of agent a for resource type r
$B(a,t) : \mathcal{A}^t \times T \rightarrow O$	The decision function which determines the behavior of agent a at time t
$O = \{o_1, \ldots, o_n\}$	The set of n possible actions
$P_f(a, s^t_a, r) : \mathcal{A}^t \times T \times R \rightarrow [0,1]$	The probability at time t of agent a to forage resources of type r
$P_r(a,t,c_r) : \mathcal{A}^t \times T \rightarrow [0,1]$	The probability of reproduction of agent a at time t, capped at c_r
c_r	The normalization constant of reproduction
$P_d(a,t,c_d) : \mathcal{A}^t \times T \rightarrow [0,1]$	The probability of death of agent a at time t, capped at c_d
$d(a,t) : \mathcal{A}^t \times T \rightarrow \mathbb{N}_{0,\leq L}$	The age function, linearly increasing in time.
c_d	The normalization constant of death
$\mathcal{I}^t_a = \{i \in E^t : i \text{ is visible to } a\}$	The perception vector of agent a at time t
$H^t_{a,r} = \sum_{t \geq j \in T_{a,r}} g(a, s^j_a, r)$	The foraging history of agent a and resource type r at time t
$T_{a,r} = \{t \in T$: a choses to eat $r\}$	The times at which agent a executes a foraging action on a resource of type r
$L \in \mathbb{N}_{>0}$	The simulation length
$l \in \mathbb{N}_{>0}$	The length of seasons
$H^t_a = \sum_{r \in R} H^t_{a,r}$	The foraging history of agent a at time t
$b : \mathcal{I} \rightarrow \mathbb{R}^n$	The behavior function which assigns a value to every action

https://doi.org/10.1371/journal.pone.0219502.t001

2.2 Agents

The agents serve as an abstract model for simple biological entities, which require to find food and forage in order to survive and reproduce. Agents are able to perceive their surroundings, i.e., defined as their range one Moore neighborhood, in the grid-like environment; the perception vector is denoted as $\mathcal{I}$. A range one Moore neighborhood in a two-dimensional square grid is comprised of eight surrounding cells (horizontal (2), vertical (2) and two diagonals(4)).

Agent actions can either be a movement, that displaces them by one cell in the environment, or foraging, that consumes any available food in their current location. A foraging action fails if the current cell does not contain any resource, or randomly with probability $1 - P_f(a, s^t_a, r)$ otherwise (for agent a with skill level s at time t for resource type r).

2.2.1 Aptitude and skill. The foraging strategy of an agent is determined by two parameters: (i) *aptitude*, which defines the value encoded in the genome and inherited from the parent, and (ii) *skill*, which defines the corresponding phenotypic expression and models the trade-off of specialization in a specific resource type [30] by influencing the probability of successful foraging.

For this reason, the skill of an agent is a determinant factor for the energy intake, their ability to reproduce, and consequently the fitness of the agents. The aptitude remains constant during the whole lifetime of an individual and changes only between generations via random mutations during reproduction. The initial value of skill at birth is determined by the inherited value of aptitude. If the skill parameter is plastic, i.e. adapts to the environment during the agent's lifetime, then the value of aptitude influences only indirectly the energy intake of learning agents.

2.2.2 Energy level. The energy level of an individual depends on three factors: (i) the availability of resources in the environment at each given time, (ii) the individual skill which determines the probability of successful foraging, and (iii) the individual behavior which determines what actions to execute for a given configuration of the environment.

More formally, the fitness function $f(a, t)$ of an agent $a \in \mathcal{A}$ at time t is defined as the total energy intake:

$$f(a,t) = \sum_{t \in T} \epsilon H_a^t \tag{1}$$

Where H_a^t is the foraging history and ϵ is the energy level increase factor.

Fitness depends on the foraging success function g:

$$g(a, s_a^t, r) = \begin{cases} 1 & \text{with probability } P_f(a, s_a^t, r) \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

2.2.3 Foraging, reproduction, death, and fitness. The experimental design introduces a trade-off between the foraging success of different resource types, determined by the skill s_a^t: agents can either become generalists, i.e. be able to forage both resources with a low probability, or specialize, i.e. be able to forage one resource with a high probability and lose the ability to forage the other.

Successful foraging increases the energy of an individual which determines the probability of reproduction. As agents compete for the same limited resources, efficient foraging translates to high reproduction rate.

The probabilities of foraging P_f, reproduction P_r and of death P_d are defined as:

$$\begin{aligned} P_f(a, s_a^t) &= (s_a^t)^q \\ P_r(a, t, c_r) &= f(a,t)/c_r \text{ if } f(a,t) < c_r \text{ else } 1 \\ P_d(a, t, c_d) &= d(a,t)/c_d \text{ if } d(a,t) < c_r \text{ else } 1 \end{aligned} \tag{3}$$

With a linear relation between skill and probability of foraging success, i.e. $q = 1$, the average total intake of an agent is equivalent to the average resource distribution: a specialist agent forages with certainty one type of resources but none of the other, while a generalist agent forages each resource with 50% probability. Assuming a non-linear relation between skill and foraging probability instead, i.e. $q > 1$, then a specialization leads to higher fitness than a generalization.

The effects of these values can be found in the supplementary material, S3–S5 Figs.

The framework determines the reproduction and death events by means of a genetic operator called *roulette wheel selection with stochastic acceptance* (as in Torney et al. 2011 [32]), according to which agents reproduce asexually with a probability P_r proportional to their fitness and die with a probability P_d proportional to their age. Upon reproduction, the energy

level ϵ of the parent is split equally between the parent and the offspring and the offspring inherits a randomly-mutated copy of the parent's genetic configuration.

2.3 Agent behavior

An agent's desired behavior $B(a, t) = \arg\ \max_{o \in O}(b(\mathcal{I}_a^t))$ associates the desired action to each perception vector $\mathcal{I}$, containing a representation of the surroundings that informs about the location and presence of resources. This mapping between perception and action can be achieved by different techniques, e.g. an artificial neural network. The success of the desired action is determined by the *skill* value, which is defined as the phenotypic expression of the *aptitude* genotype.

The aptitude and the mapping $B(a, t)$ changes from one generation to the next due to random mutations, and learning allows the inherited skill and the phenotypic expression of the mapping $B(a, t)$ to be more suited to the current state of the environment.

2.3.1 Agent types and learning. Two types of agents are introduced: reactive agents keep their behavior and skill constant throughout their lifetime, as they are a direct expression of the genotype, while learning agents adapt their behavior and skill according to their experience via reinforcement learning [16, 25, 26, 33–35]. Learning optimizes the expected reward associated with successfully foraging a resource of any type. Different reinforcement learning architectures are evaluated: Q-Learning [36], reinforcement learning based on a Restricted Boltzman Machine [37], Deep Reinforcement Learning [38] and reinforcement learning based on a single feed forward perceptron. The results presented in the main text are based on a single feed forward perceptron, see the supplementary material for further details, Section B in S1 File).

If learning is disabled (reactive agents), weights and skills cannot be learned and remain constant and equal to the inherited value for the whole lifetime of the individual, hence individuals are selected based on their inherited aptitude value. If learning is enabled, the behavior can adapt to changes in the environment. Specifically, the adaptation process happens through directly increasing the skill value after every successful foraging event and by using the successful foraging event as a reward signal in the reinforcement learning algorithm.

2.3.2 Genotype and mutations. Upon reproduction, an offspring is generated that contains a mutated copy of the parent's genome, consisting of the initial weights of the neural network, prior to any learning, and an additional gene called *aptitude*. These values are used to initialize the phenotype of the offspring.

2.3.3 Reinforcement learning. Although, modeling biologically realistic entities is outside the scope of this paper; the study of the biological feasibility of different learning techniques including different versions of reinforcement learning, have shown that reinforcement learning is being able to reproduce certain human decision-making process and equilibrium [39–41]. More recently it has been shown that human level strategies can arise from reinforcement learning-based systems, even without human data [42]. In the case of this implementation of reinforcement learning, the behavior function of an agent a takes the form of $b(\mathcal{I}_a^t) = Q(\cdot, \mathcal{I}_a^t)$ which indicates the Q-values for all actions and state $\mathcal{I}_a^t$. The mapping between perceptions and actions is done via a neural network. Agents perceive their the environment, specifically, they are able to see a subset of the grid centered at their location (range one Moore neighborhood) and are able to identify food sources within this visual range, $\mathcal{I}$. For the current model, a 3×3 region is observable and the food sources are observable but without the specificity of the amount of food contained. Based on this perception and using the neural network based choice model agents chose an action from their action space: move (north, south, east, west) or eat. Agents with different learning algorithms (Neural Network type) behave differently when

faced with a variable environment, in terms of convergence and *adaptation to change* (see the supplementary material, Sections B and C in S1 File).

The agent skill is learned by increasing its value by ΔS after every time it performs foraging successfully, while for the choice of action the learning algorithms are based on the Reinforcement Learning approach, Q-Learning [36]. The Q-Table, a mapping from states/perceptions $\mathcal{I}$ and possible actions O to the quality value of each action for that state $Q(\mathcal{I}, O)$, of the original Q-Learning approach is replaced by a Q-Network as per [38] and the corresponding algorithm for the specific Q-Network structure is used for its training. The following equation describes the update to the quality values:

$$\Delta Q = \left(\overbrace{\underbrace{r_{t-1}}_{\text{reward}} + \underbrace{\gamma}_{\text{discount factor}} \cdot \max_{O} Q(\mathcal{I}_t, O)}^{\text{learned value}} - \underbrace{Q(\mathcal{I}_{t-1}, O_{t-1})}_{\text{old value}} \right) \tag{4}$$

$$\underbrace{Q(\mathcal{I}_{t-1}, O_{t-1})}_{\text{new desired value}} \leftarrow \underbrace{Q(\mathcal{I}_{t-1}, O_{t-1})}_{\text{old value}} + \underbrace{\alpha^{rlearn}}_{\text{learning rate}} \cdot \Delta Q \tag{5}$$

The results presented in the main text are based on Reinforcement learning using a single layer feed forward perceptron as its network architecture to "store" and query the Q-values, trained with backpropagation (PQL). The Q-network structure is $b(\mathcal{I}) = W \cdot \mathcal{I} + \beta$ where $\mathcal{I}$ is an input vector, W are the weights of the neural network and β the biases associated to the input layer. Further details about the variations depending on different neural network structures can be found in the supplementary material, Sections B and C in S1 File.

2.4 Measures

The degree of specialization of a population is measured with different metrics:

(I) the distribution of individual aptitudes across the population, according to which a higher frequency of extreme values corresponds to a more specialized population, (II) the individual foraging history, i.e. the frequency of successful foraging actions for a specific resource type, according to which extreme values indicate a specialized diet, (III) standard measures of group behavior that quantify the rate of consumption of resources ([43], page 241).

The degree of specialization of the population is measured by the distribution of aptitudes (I) at each given timestep, normalized by the population size at that timestep:

$$M^{I}(v, t) = |\{a \in \mathcal{A}^t \ : \ s_a^t = v\}| / |\mathcal{A}^t| \tag{6}$$

The foraging history (II) of the population at value x is measured as the frequency of individuals in the population who, during their lifetime, foraged a specific proportion of type r resources corresponding to x:

$$M^{II}(x, r) = |\{a \in \mathcal{A} \ : \ H_{a,r}^L / H_a^L = x\}| / |\mathcal{A}| \tag{7}$$

Additionally, standard measures of group behavior (III), taken from [43], page 241, are used to quantify the specialization of the population. The measures are defined and explained in the supplementary materials, Section H in S1 File.

While (I) measures the characteristics of the genotype, (II) and (III) measure the behavior of the agents which is determined by the phenotype.

3 Results: Computational model

3.1 The Baldwin effect

Previous work in the literature about the Baldwin effect found that the evolutionary process can be either speed up or slowed down [2] depending on the learning mechanism, the fitness function and the starting conditions of the population. Simulations are performed to verify whether or not the Baldwin effect exists in a cyclical environment, a question that, to the best of our knowledge, has not been answered before [17].

The existence of the Baldwin effect is evaluated by means of simulation by comparing the speed of genetic assimilation of phenotypic features as a function of the learning ability.

Fig 1 shows a comparison over time of three agent types in terms of the genetic assimilation of aptitude values due to changes in skill value:

- Reactive agents: baseline, i.e. unable to learn.
- Learning (Actions): agents that can modify their own actions through learning, a speedup in the genetic assimilation is observed.
- Learning (Actions & Skill): agents that can modify their own actions and their skill through learning, a slowdown in the genetic assimilation is observed.

The only difference between agent types concerns what traits can be learned. All other parameters of the learning algorithms are constant across types. The dependence of the speed of genetic assimilation on the degree of learning confirms the presence of the Baldwin effect.

3.2 A new effect: The Baldwin veering effect

This experiment investigates whether the Baldwin veering effect exists, i.e. a trait evolves by effect of plasticity that does not correspond to the plastic response induced by a cyclically-changing environment.

Slowly-changing environments allow populations to adapt via natural selection. Learning helps natural selection traversing the space of genetic configurations [44], and does so on a shorter timescale, therefore learning might speed up or delay this process. In quickly-changing environments, which change faster than the evolutionary timescale, learning and natural selection take on two different roles: Learning improves the behavior of agents in response to environmental variability, while natural selection improves the efficiency of learning.

The Baldwin veering effect is present if the following two conditions are verified: (i) the evolved trait differs from the environment-induced plastic response, i.e. genome and the phenotype converge towards different values, and (ii) this effect is determined by the presence of plasticity, i.e. the genetic configuration evolved by learning agents differ from that evolved under the same conditions by reactive agents.

Two genetic configurations are considered: a *specialist configuration* is defined as a genome whose aptitude evolves to one of the extreme values, i.e. specializes in either resource type, *generalist configuration* is defined as a genome whose aptitude evolves to an intermediate value. Different genetic configurations correspond to different initial learning efforts in terms of time required to adapt to the environment; assuming that an individual has the same probability of being born in either season, the optimal genetic configuration should reduce equally the effort of learning either skill.

The first condition is verified in Fig 2 by comparing the evolution over time of the inherited aptitude of populations of learning agents in a slowly-changing (Left) and in a quickly-changing environment (Right). Being able to quickly adapt to changes in the environment, the phenotype of learning agents tracks changes in resource availability, hence the induced

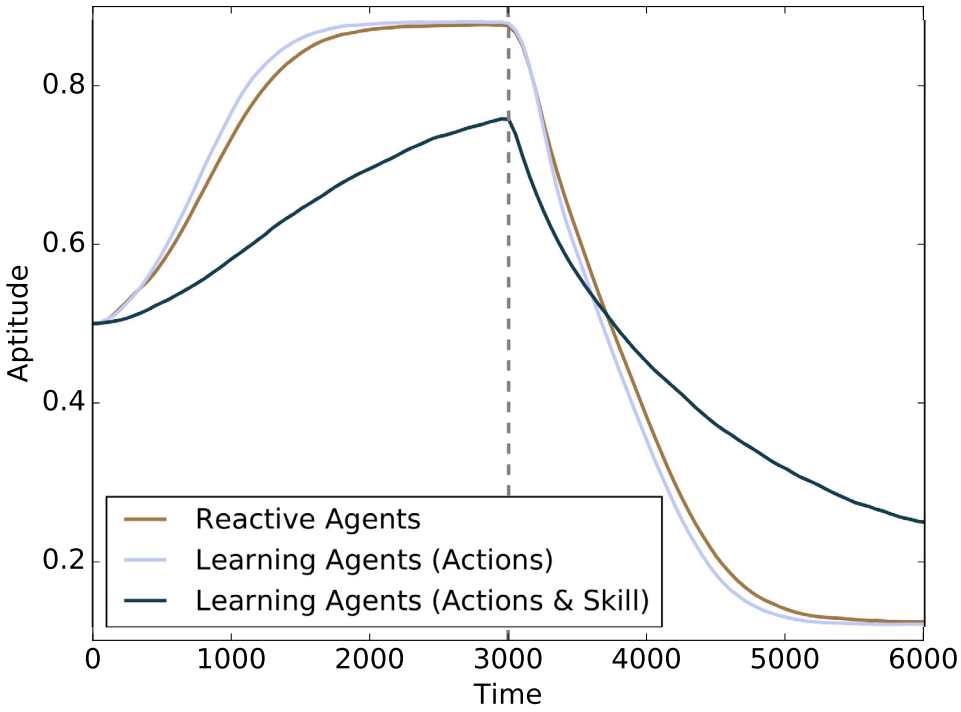

Fig 1. The Baldwin effect: Speed of genetic assimilation (genotype). Genetic assimilation of aptitude over time in two different learning populations, compared to a baseline of reactive agents. The simulations show that the speed of genetic assimilation changes with respect to the baseline, depending on the configuration of the learning algorithm, demonstrating that the Baldwin effect can speed up or slow down the genetic assimilation of aptitude. The dashed vertical line indicates a change of season, i.e. resource availability. Confidence intervals at the 95% confidence level are not shown as their size is negligible.

https://doi.org/10.1371/journal.pone.0219502.g001

plastic response is a specialized strategy corresponding to the most abundant resource type. In a slowly-changing environment, the genetic trait evolves towards the induced plastic response, while in a quickly-changing environment, the genetic trait evolves an intermediate value that corresponds to a generalist strategy which is not induced at the phenotypic level.

The second condition is verified in Fig 3 by comparing the evolution over time of the inherited aptitude of a population of reactive agents (Left) with that of a population of learning agents (Right). Both populations are initialized with an intermediate aptitude value, which evolves over time until it converges to some configuration after around 4000 timesteps. Reactive agents evolve extreme aptitude values, i.e. a specialist configuration. Specifically, half of the population evolves a high aptitude value (specialist in one type of resource) and the other half a low aptitude value (specialist in the other type of resource). The foraging success of reactive agents is determined directly by the static skill as inherited from the aptitude value, hence each half of the population specializes in foraging one or the other type of resource. For learning agents the foraging success is determined by the skill level, whose initial adaptation effectiveness is determined by the aptitude value. As a consequence, learning agents evolve an intermediate aptitude value, i.e. a generalist configuration, which allows them to adapt quickly to any environmental condition. Fig 4 highlights the difference between genetic configurations evolved by the two populations at the end of the simulation.

In the following section, we present further supporting evidence for these results.

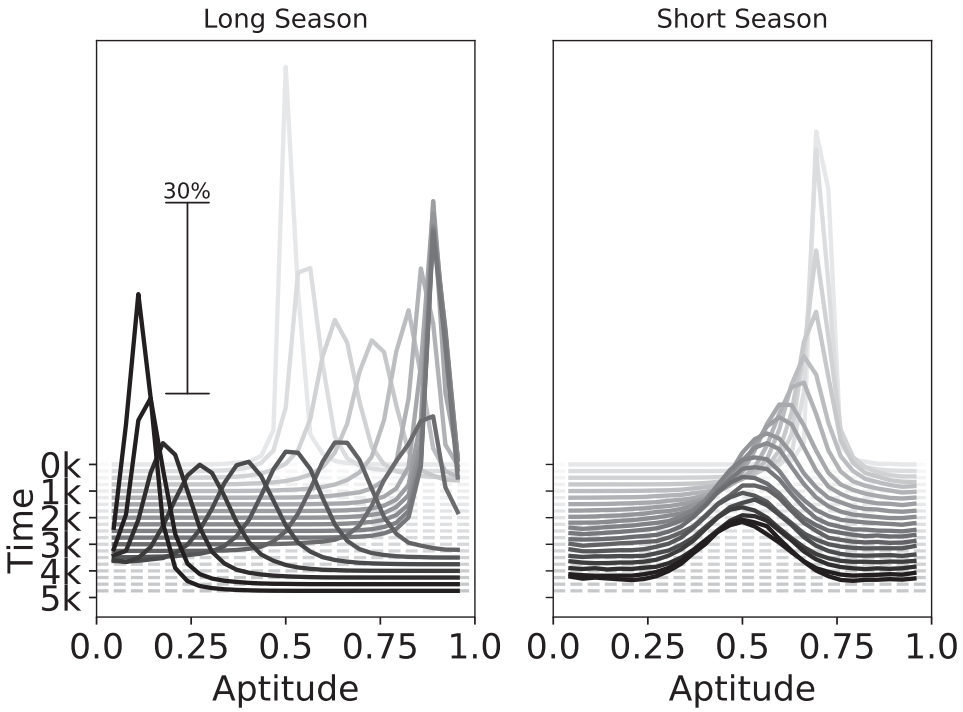

Fig 2. Comparison of aptitude distributions (genotype) of plastic agents in environments with different season length. The graphs show the change in the distribution of aptitude values across genomes of individuals in the population (horizontal-axis, Aptitude) over the course of the simulation (depth-axis, Time). Each point in the graph represents the frequency at which a specific aptitude value was present in the population (vertical-axis, % of the population) at a given time. The plot shows plastic agents during a long season (left) and short season (right) scenarios. During long seasons the population tracks the plastic induced response, while during short seasons the genetic trait evolves towards an intermediate value distribution. "k" denotes thousand abstract time-units.

https://doi.org/10.1371/journal.pone.0219502.g002

3.3 Differences in individual behaviors

In order to verify that a difference in genetic configuration actually results in different behaviors, in this section, we analyze reactive agents instantiated with the genetic configurations of the agents that are alive during the last time-step of the previous simulations (see Fig 4). In order to produce a fair comparison, reproduction is also disabled, this way the only variable in the simulations is the genetic configuration which remains constant during these simulations and is expressed directly in the phenotype.

The goal of these new simulation set is to quantify the difference between genetic configurations evolved by different populations, this is achieved by evaluating the behavior that such configurations encode.

In these new simulations, the environment is set to have only one season and contains an equivalent quantity of both types of resources. An abundance of both resource types allows any foraging strategy to perform at its best, hence contributing to a fair comparison of different foraging strategies in terms of foraging success. The behavior of individuals is compared with the measures of foraging history and of group behavior, which are described in Section 2.4.

Fig 5 shows the foraging history of the two populations of study. Additionally, it shows the foraging history of 2 baseline populations. These baseline populations are also reactive agents instantiated with genetic configurations specifically aimed to produce specialist behavior (being able to eat only one food type with high probability) and generalist behavior (being able to eat both food types with 50% probability). The foraging history shows that the behaviors in

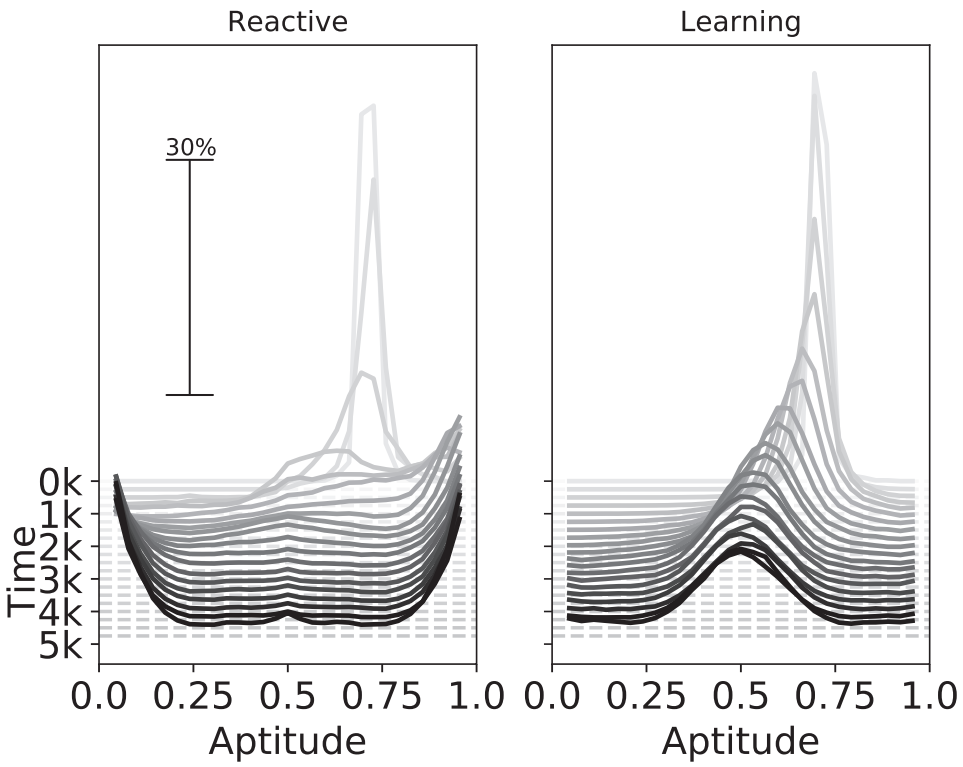

Fig 3. Comparison of aptitude distributions (genotype) between reactive and plastic agents in environments with short seasons. The graphs show the change in the distribution of aptitude values across genomes of individuals in the population (horizontal-axis, Aptitude) over the course of the simulation (depth-axis, Time). Each point in the graph represents the frequency at which a specific aptitude value was present in the population (vertical-axis, % of the population) at a given time. The left plot shows a population of reactive agents while the right plot a population of plastic agents, the populations evolve two different distributions, confirming that plasticity/learning can change the outcome of evolution. "k" denotes thousand abstract time-units.

https://doi.org/10.1371/journal.pone.0219502.g003

the two populations of study differ (cf. Fig 5), namely the population instantiated with the last reactive configuration is split into two groups of comparable size, each of which is specialized in foraging one type of resource, while the population instantiated with the last learning configuration has a more uniform foraging pattern which includes more generalists. The measure of individual foraging history is quantified by the frequency of foraging resources of type one, e.g. a value of 90% indicates that 90% of all resources foraged by the agent were of type one, and the remaining 10% of type two. These values are then aggregated across the population to determine the frequency of different values of foraging history. The inset of Fig 5 reports the L_2 Norm between the distributions; learning configuration agents distribution is closer to the generalists' distribution (0.19) while reactive configuration agents distribution is closer to the specialist distribution (0.25).

Besides the measure of foraging history, different standard measures of group behavior [43] are used to compare the behavior of the populations (cf Fig 6). The interpretation of these measures is not straightforward, so baselines are added for reference: the dashed line represents the value of a population where half of the agents specialize in one resource and the other half in the other resource, while the continuous line represents a population of generalists.

The measures confirm that the learning configuration agents develop a generalist foraging strategy, both on the group level (among-resource diversity) and on the individual level (within-individual diversity). In contrast, last reactive configuration agents develop a more

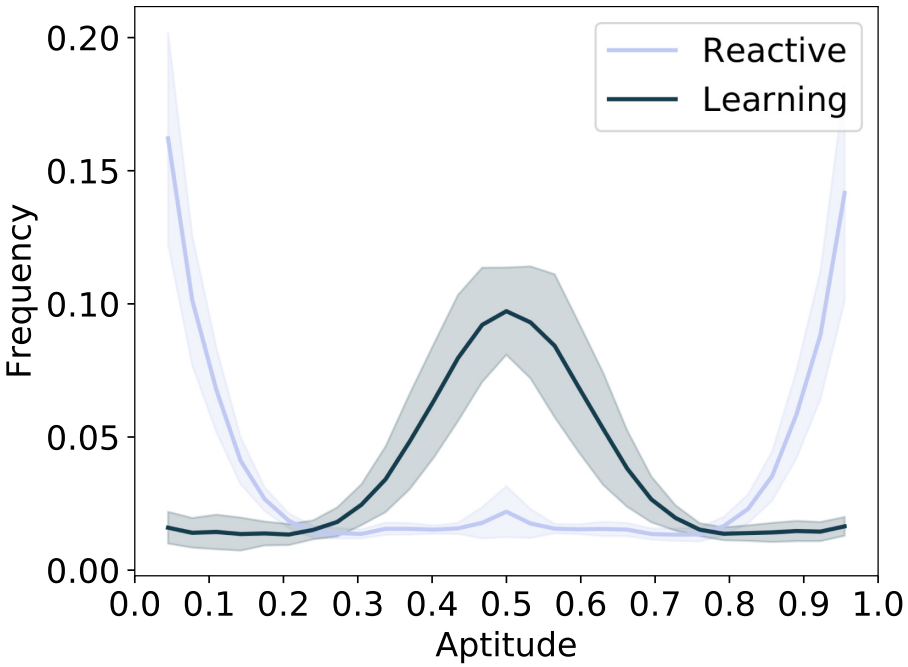

Fig 4. Comparison of final aptitude distribution (genotype) between reactive and plastic agents in environments with short seasons. The graphs show the average distribution of aptitude values (x-axis) across genomes of individuals in the population for the last 1000 iterations of the simulation, shaded areas represent confidence intervals at the 95% confidence level. The lines indicate the frequency at which a given aptitude occurs in the population. Plastic agents evolve a different distribution than reactive agents, confirming that learning can change the configuration to which evolution converges.

https://doi.org/10.1371/journal.pone.0219502.g004

specialized foraging strategy on the group level (among-resource diversity). Understanding whether or not reactive configuration agents develop a specialized foraging strategy on the individual level is not straightforward, as a high value of among-resource diversity can either mean that different agents have different specialized diets or that agents have generalized diets. Combining this measure with that of within-individual diversity, which indicates a specialized diet on the individual level, allows us to conclude that specialization occurs also on the group level.

4 Results: Analytical model

The results outlined in the previous section showcase the existence of the Baldwin veering effect, but give little information about the process behind it. This section introduces and analyzes the predictions of an analytical model, inspired by previous work [45], which gives possible explanations to the simulation results and identify the conditions under which the Baldwin veering effect manifests. The model defines a fitness function for a generic individual, the evolutionary process is not explicitly modeled so evolutionary outcomes are inferred from considerations about the relative fitness of different individuals. Time and location of agents are not explicitly modeled, this abstraction is sensible because of the deterministic nature of seasonal changes, i.e. the environment displays the same conditions on average over each seasonal cycle. More fine-grained results about evolution and its dynamics might be obtained by pairing the fitness function with an existing model of evolution, e.g. [45, 46], such effort is outside the scope of this paper and is left for future work.

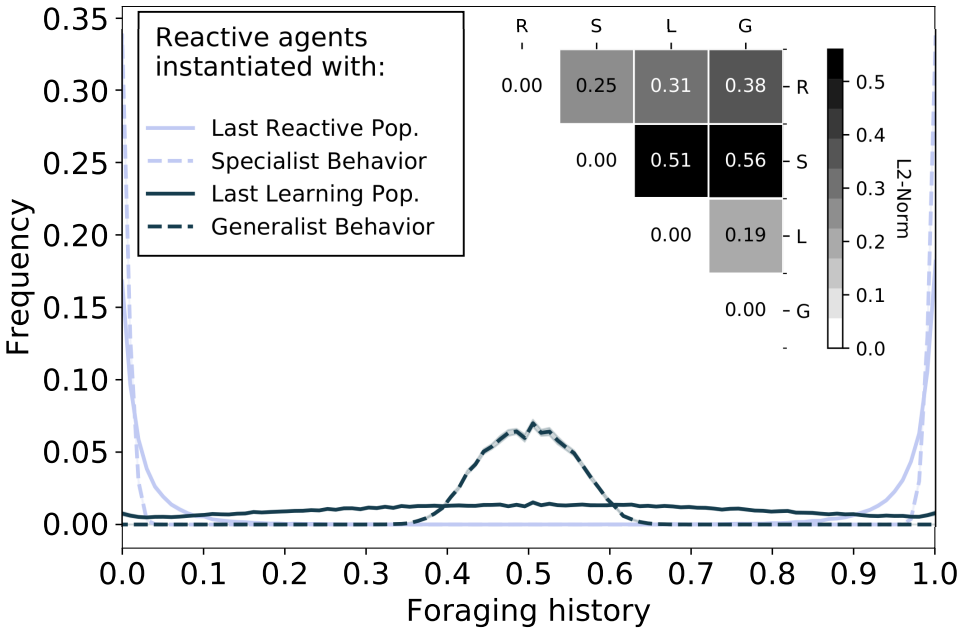

Fig 5. Comparison of foraging history (phenotype). The plots show the frequency at which a given value of foraging history occurs in the population. Foraging history is computed as the percentage of successful foraging actions or resources of type 0. A frequency of 0.2 associated to a value of foraging history of 0.4 means that 20% of individuals in the population foraged during their lifetime 40% of the time resources of type 0 and 60% of the time resources of type 1. In this simulation, all agents are reactive without the ability to reproduce and their genetic configuration is initialized from the distribution obtained from the second experiment (see Fig 4). Distributions of foraging actions resemble the distributions of aptitudes, confirming that different aptitude distributions produce different behaviors. Dashed lines represent baseline populations, where all agents have an aptitude value of 0.5 (generalist configuration) or half of the population has an aptitude value of 0.05 and the other half 0.95 (specialist configuration). Shaded areas (of negligible size) represent confidence intervals at the 95% confidence level. Inset is the L_2-Norm between the different distributions.

https://doi.org/10.1371/journal.pone.0219502.g005

4.1 Description of the analytical model

The environment contains two types of resources, $j = \{0, 1\}$, whose proportion is denoted by π_0 and π_1.

The fitness W_i of a reactive agent i is formulated as follows:

$$W_i = \pi_0 \cdot r_{i,0} + \pi_1 \cdot r_{i,1} = \pi_0 \cdot s_{i,0}^q + \pi_1 \cdot s_{i,1}^q \tag{8}$$

Where the foraging success $r_{i,j} = s_{i,j}^q$ is determined by the agent's skill $s_{i,j} \in [0, 1]$ (which is equal to the aptitude level, being it a reactive agent) and by a parameter $q \in \mathbb{N}_{\geq 0}$ which defines the relation between skill and foraging success. If the parameter $q = 1$, specializing on one resource and generalizing on two resources lead to the same foraging success. If $q < 1$ generalization becomes more beneficial than specialization as intermediate aptitudes produce a higher foraging success than extreme ones. Vice versa, specialization is more beneficial when $q > 1$ as the reward function is concave, a requirement for the co-existence of specialists and generalists in the same environment [47].

Following the design of the computational model, the two skills of an agent, as well as the resource proportions, are assumed to be complementary, i.e. $s_{i,0} + s_{i,1} = 1$, $\pi_1 + \pi_0 = 1$, therefore

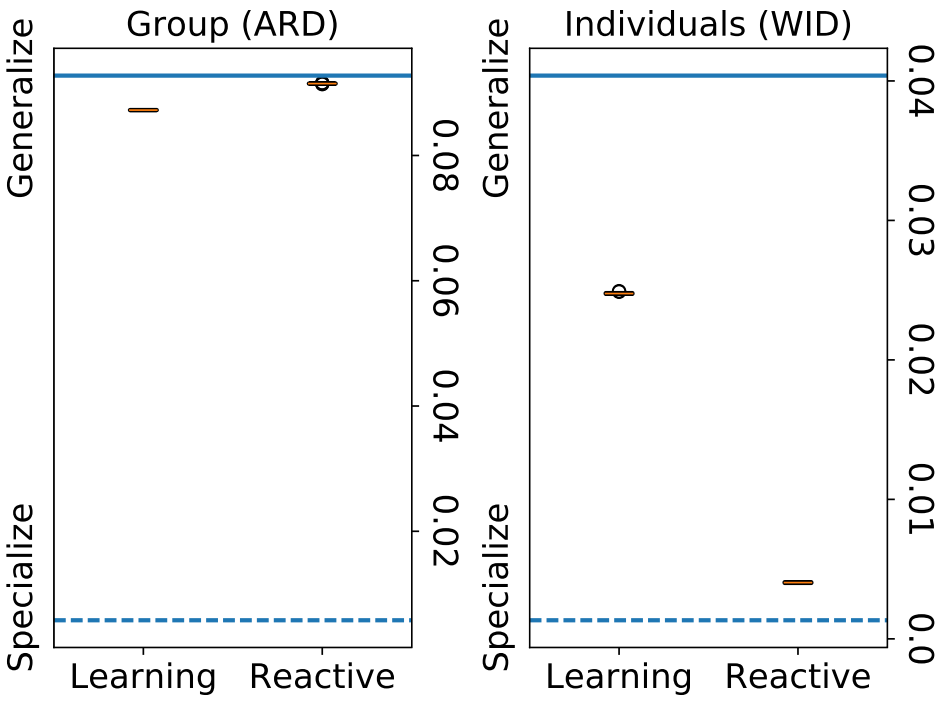

Fig 6. Measures quantifying the behavior of the population (phenotype). Left: among-resource diversity (ARD) quantifies the behavior of the population, both populations display a similar generalist behavior. Right: within-individual diversity (WID) quantifies the behavior of individual agents, learning agents behave more generalist than reactive agents. In this simulation, all agents are reactive without the ability to reproduce and their genetic configuration is initialized from the distribution obtained from the second experiment (see Fig 4). These results confirm that a difference in aptitude distribution corresponds to an actual difference in behavior. The solid line represents a baseline population in which all agents have an aptitude value of 0.5, the dashed line represents a baseline population in which each half of the agents have an aptitude value of 0.05 and 0.95 respectively.

https://doi.org/10.1371/journal.pone.0219502.g006

the notation can be simplified by defining $s_i \coloneqq s_{i,0}$, $1 - s_i \coloneqq s_{i,1}$ and $\pi_1 = 1 - \pi_0$ which leads to:

$$W_i = \pi_0 \cdot s_i^q + (1 - \pi_0) \cdot (1 - s_i)^q \tag{9}$$

In order to model learning agents, a new parameter δ is introduced which represents plasticity. The parameter c determines the cost of plasticity [48, 49]. A learning agent is not constrained by its inherited aptitude α, as its skill can adapt to changes in the environment. The value of δ determines the skills an agent can express by defining the maximum and minimum skill values: this range is centered on the aptitude and spans in both directions (cf. Fig 7), $s_i = \alpha_i \pm \delta$. Given that the skill value is limited in the domain [0, 1], the previous expression for the bounds of skill values s is limited as follows, $s_i = min(1, \alpha_i + \delta)$ for the beneficial side and $s_i = max(0, \alpha_i - \delta)$ for the dis-favorable effect. For example an individual with aptitude 0.1 and $\delta = 0.6$ can express any skill value in the range [0, 0.7]. As the aptitude is also constrained to the range [0, 1] the range of meaningful δ is also between [0, 1]. In this model, we consider only the effect of plasticity that increases the skill level (i.e learning that improves a skill):

$$W_i = \pi_0 \cdot min(1, (\alpha_i + \delta))^q + (1 - \pi_0) \cdot min(1, (1 - \alpha_i + \delta))^q - c \cdot \delta \tag{10}$$

For simplicity, the model assumes that agents adapt instantaneously to the environment by adopting the best available skill value *for each resource type*, i.e. skill of $s_i = \alpha_i + \delta$ for resource type π_0 and skill of $s_{i,1} = \alpha_{i,1} + \delta = 1 - (\alpha_i - \delta)$ for resource type π_1, which maximize the fitness

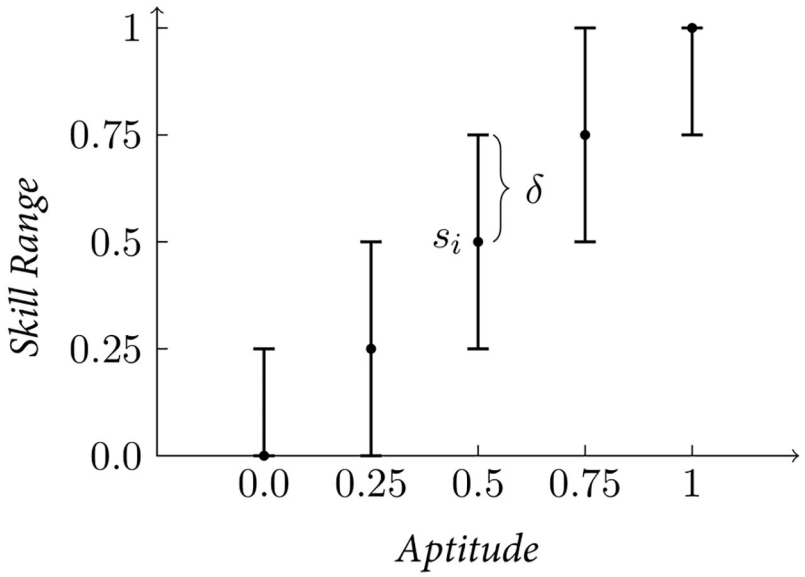

Fig 7. A sketch explaining the skill range δ. Skill ranges obtained with a fixed value of δ and different aptitudes.

https://doi.org/10.1371/journal.pone.0219502.g007

function. The speed of learning, also called time lag, is modeled by reducing the value of δ (cf. Fig 8). In practice the value of δ depends on the ratio between the speed of learning and the season length: a slower learning mechanism reduces the distance to which the value can change, similarly, a shorter season reduces the number of experiences an agent has during a season.

4.2 Analysis: Baldwin veering effect

Fig 9 shows how different aptitudes compare, in terms of fitness, for varying values of plasticity δ. A combination of aptitude and plasticity associated with a higher fitness value produces more fit individuals that are favored by natural selection. The red circles represent the globally optimum aptitudes for a given value of δ, i.e. the attractors in genetic configuration space of

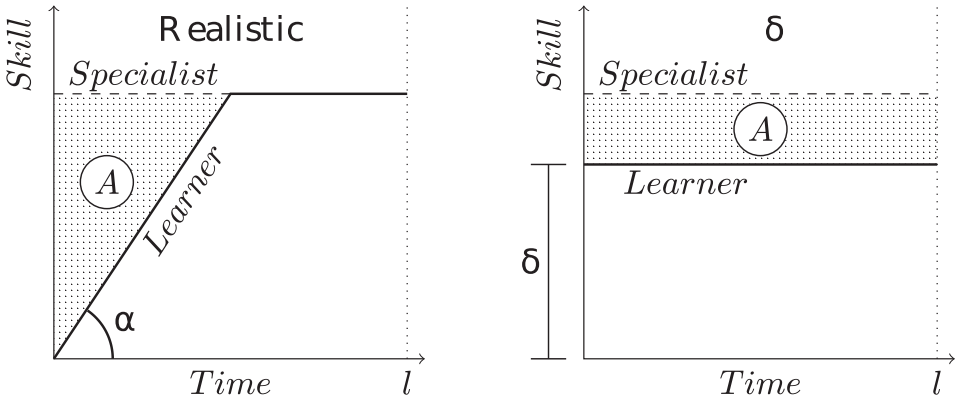

Fig 8. A sketch of modeling assumptions. The graph shows the change in skill level over time of a hypothetical learning individual. The shaded area represents the cost of adaptation: the loss in fitness caused by adapting to the environment with respect to an already adapted individual (specialist). Learning requires time to adapt, defined by the speed of learning α. This delay is modeled by reducing the plasticity δ such that the size of area A is the same.

https://doi.org/10.1371/journal.pone.0219502.g008

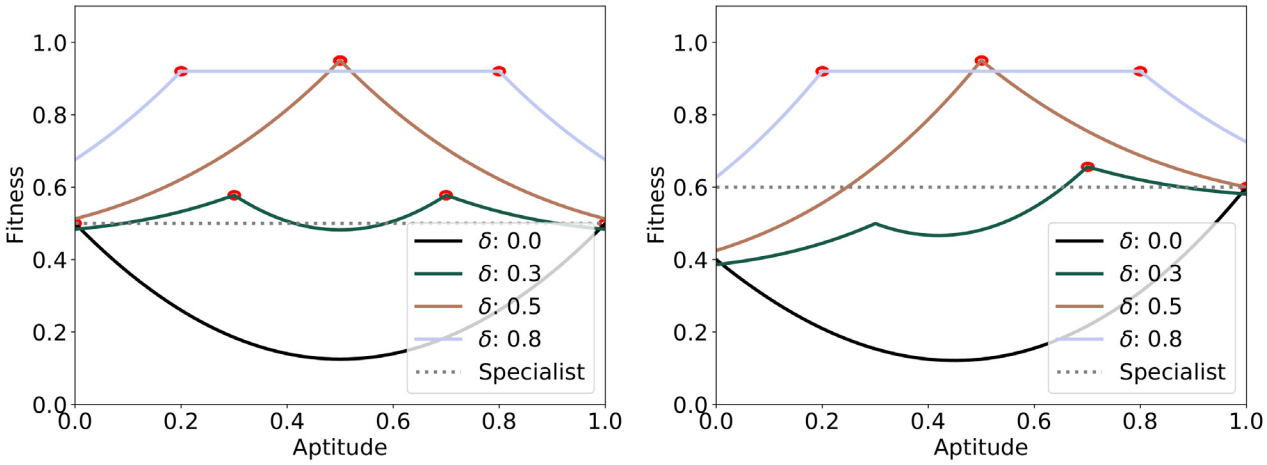

Fig 9. Fitness corresponding to different combinations of aptitude and δ, for $q > 1$ and $c > 0$. The plot shows the fitness predicted by the analytical model for given values of aptitude and δ. Left: $\alpha_0 = 0.5$, right: $\alpha_0 = 0.6$. The red circles represent the maximum fitness achievable for a given value of δ, i.e. the attractors of the evolutionary process. Increasing the value of δ from 0 to 1, the optimal aptitude values start at the extremes (0,1) and move towards the center as δ increases. The maximum fitness is obtained for $\delta = 0.5$, where there is only one maximum for an aptitude of 0.5. For values of $\delta > 0.5$ a range of aptitudes, centered on and expanding from 0.5, maximizes the fitness. The dotted line corresponds to the fitness of a specialist individual, which becomes lower than the fitness of learning individuals as values of δ increase. Also note that the introduction of learning, i.e. $\delta > 0$, changes the aptitude for which fitness is maximized, i.e. the configuration towards which evolution converges.

https://doi.org/10.1371/journal.pone.0219502.g009

the evolutionary process. If $\delta < 0.5$ agents evolve a specialist configuration, as opposed to a generalist configuration if $\delta = 0.5$. Note that the configuration with $\delta = 0.5$ and aptitude $\alpha_i = 0.5$ maximizes the fitness as it allows agents to choose any skill value in the range [0, 1], hence allows agents to forage both resource types with certainty. This condition is observable in the agent-based model simulation when the speed of learning is as fast as the frequency of change in the environment, i.e. an agent adapts its skill to a new environmental state but does so too slowly to remain specialized for a long time before the environment changes again. This confirms the existence of the "Baldwin veering effect", as any value of $\delta > 0$ changes the fitness landscape such that fitness is maximized by a different aptitude, which is then selected. These results hold even for asymmetric seasons, i.e. when the probability of one season is higher (cf. Fig 9 right).

For values of $\delta > 0.5$, learning makes an increasingly large range of aptitude values equivalent in terms of evolutionary fitness which could allow agents to generalize, but such a configuration would not evolve in reality as the overall fitness is reduced when compared to $\delta = 0.5$. These results are confirmed also for $c = 0$ and $q \leq 1$, see the supplementary material, Section G in S1 File.

Concluding, learning agents evolve an intermediate aptitude, i.e. a generalist configuration, only if learning speed is proportionate to the season length such that agents can adapt to both resource types. This result is general and holds independently of the value of q and resource proportion π_0, hence confirms that the Baldwin veering effect depends exclusively on the timescales of learning and environmental change.

5 Discussion

A common finding in the literature about the interactions between plasticity and cyclically changing environments is that plastic individuals, who can adapt to changes in the environment after a certain time lag, i.e., speed of learning, are more fit than non-plastic individuals, who are unable to adapt, when the frequency of change in the environment is faster than a certain threshold. The definition of plasticity varies in the literature: plasticity is modeled as

switching between two distinct phenotypes [14, 50], as a change in niche breadth [51], or—as in this work—as behavioral adaptation through learning [52, 53]. Related work concludes that similar patterns of specialization and generalization in the phenotype might develop also when assuming non-reversible plasticity, i.e. a phenotypical trait can assume only one specialized state in an individual's lifetime [19]. Although this work focuses on reversible plasticity, i.e. the same phenotypical trait can change from one specialized state to another, reversibility is not claimed to be a prerequisite for the existence of the Baldwin veering effect. Our claim about the existence of the Baldwin veering effect is not invalidated by whether or not the effect manifests also in the presence of non-reversible plastic traits, this is nevertheless a promising research question for future work.

It is important to note that although the concepts of specialization, i.e. adaptation to only one state, and generalization, i.e. trading-off adaptation across more than one state, are consistent across the literature, the concepts of generalist and specialist can differ substantially: while the definition of specialists adopted by this work implies the ability to specialize in only one resource, unless the agent is able to learn, other work defines them as able to specialize simultaneously on many resources [14]. To the best of our knowledge, this work is the first to investigate the effect of plasticity and cyclical changes in the environment on the evolution of the genetic configuration. Although, [54] considers local variation (cycles) the analysis is focused on a single movement to an extreme environment; our work is consistent in terms of the expected genetic assimilation. Previous work investigates the evolution of plasticity by analyzing the co-evolution of populations with different genetic configurations [55, 56], or by analyzing the scaling of plasticity itself [54] while in the current work traits are either plastic or not. Due to the differences between our models, we cannot make more extensive claims, i.e. regarding the development of co-evolving populations nor the evolution of plasticity itself. However, results pertaining to the more general aspects of the Balding effect, i.e. genetic assimilation, is consistent across the works. Other work [14] predicts that non-plastic individuals would evolve a genotype that leads to a wide tolerance function, making the individual able to adapt to a broader set of environmental configurations. This is in conflict with our finding that non-plastic individuals specialize in one environmental configuration, which leads to a split in the population. We believe this difference is caused by a modeling assumption that is relaxed in this paper, i.e. each agent can express a different phenotype (tolerance function) for each environmental state. The thread of literature looking at the evolution of artificial neural networks [52, 57] concludes that different levels of plasticity lead to the evolution of different weights. The main difference with the proposed model is that the environment changes [52] during an individual's lifetime [57], hence that model is not able to capture the effect of the frequency of environmental change on evolution which is crucial for the results presented in this work.

The results presented in this work rely on the assumption that information about the environment is always precise. Relaxing this assumption requires the consideration of imperfect perceptions. Hence, the agents need to learn an estimate of the environmental state, before they can begin phenotypic adaptation [58] or while they are adapting [19]. Previous work finds that agents with imperfect perception learn accurate estimates of the probability distribution of environmental states and demonstrates genetic assimilation of phenotypic features, i.e. the Baldwin effect [59]. These results suggest that the Baldwin veering effect does not depend on the assumption of perception accuracy.

The aim of this paper is to provide a proof of concept, not modeling realistic entities, hence the model is constrained to only two resources. Increasing the complexity of the environment, as well as introducing group behavior, is required to model any realistic ecosystem and is left for future work.

The Baldwin veering effect can be interpreted as the interaction between two different traits throughout the evolutionary process. This interpretation could be described as a co-evolutionary process between two different traits in the same genotype: (1) the evolutionary selection pressure on the existence of plasticity implies a specific evolutionary selection pressure on (2) the aptitude level.

This effect can also be interpreted as an extended form of gene interactions [18] that affects both the phenotype and the genotype.

Plastic behavior is the outcome of complex interactions between genes, for the sake of tractability, this work abstracts these interactions as the effect of one single gene called aptitude. This simplification is reasonable to model some simple natural organisms e.g. fish behavior [60] and foraging in bacteria [61]. Gene interaction has an effect on the model, i.e. the interaction between the aptitude and the gene for plasticity changes the phenotypic expression of individuals [18]. Nevertheless, the results are more than just a special case of gene interaction as the presence of a "plasticity gene" causes changes both at the phenotypic and at the genotypic level, i.e. a different genetic code evolves in the population, which in turn produces different phenotypic traits.

Although this suggests the existence of the hypothesized effect, the theory does not clarify what processes cause learning agents to evolve a generalist configuration instead of a specialist configuration. One possible mechanism is that a generalist configuration allows individuals to have a more constant foraging success than a specialist configuration, as a skill level that oscillates around an average value allows individuals to forage more or less constantly throughout their lifetime, while a skill level oscillating around any of the extremes would result in periods of high and periods of low foraging success. An imbalance in foraging success translates to higher variance in offspring number, which is known to reduce the fitness [62]. Another possible mechanism for the evolution of a generalist configuration would be to provide indirect rewards: then, an intermediate aptitude would increase the evolutionary fitness indirectly by allowing for a faster adaptation to any environmental configuration. A similar mechanism has been described in the literature about intrinsic motivation, where evolution favors actions that are providing rewards only indirectly. For example, it has been found that curiosity and playfulness at a young age can improve fitness at a later age [63]. Understanding what processes cause the evolution of a generalist configuration is a worthy result in its own right which should be addressed in future work.

Another relevant observation is one of the convergent phenotypic outcomes. This phenomenon is highlighted in the fact that the neural network weights defining the desired behavior are initialized randomly and individuals with different genetic composition (weights) converge to similar behaviors but different weight composition. This shows that the large space of weights combinations posses a large number of equivalent optimums. Further studies in this topic might provide deeper insights relevant to the machine learning community.

Future work will also verify the predictions of the analytical model within the agent-based simulation framework, in particular, that there exists a configuration for which a learning population splits into two groups of specialists with aptitude values in [0, 0.5] and [0.5, 1] respectively, and a configuration in which learning population evolve a uniform distribution of aptitude values.

6 Conclusions

Plasticity, e.g. learning, is known to influence the speed at which evolution converges to some "target" configuration. This work, in contrast, addresses the question of whether or not plasticity in a cyclically changing environment can lead to the evolution of a different genetic

configuration. Following previous work, this question is answered by means of an agent-based model of a foraging task, with cyclical variability in the resource distribution. Additionally, this result is confirmed through an analytical model.

Experimental and analytical results show the existence of the Baldwin effect in a cyclical environment and identify the novel "Baldwin veering effect", i.e. a trait (generalist configuration) evolves by effect of plasticity that does not correspond to the plastic response induced by a cyclically-changing environment (specialist configuration) and the conditions under which it exists. A mathematical model verifies that the introduction of plasticity in the phenotype changes the fitness landscape in a way such that a generalist configuration becomes the global optimum in the space of genotypes.

These results are relevant for the literature of Evolutionary Biology, as they expand the understanding of how phenotypic plasticity influences evolution and present a novel effect caused by the interaction between learning and evolution. These results might also help to understand the effect of a fast learning process on a slow learning process in another context, which has a cyclical component, for example, opinion formation in settings where learning [64, 65] mediates the rate of exposure to different opinions [66, 67].

6.1 Data availability

The source code used to generate and analyze the data-sets is available on GitHub [68, 69]. Other information, including the parameters and libraries used, is provided in the Supplementary information, see Sections D and E in S1 File.

Supporting information

S1 Fig. Comparison of different learning algorithms. Each graph represents the frequency over time of an agent choosing to forage each resource type whenever the corresponding resource is available. A higher value produces a higher fitness, assuming the corresponding resource is available in the environment. Each curve is the average of 300 independent simulations. Season length is 3000 and all simulations start in the same season.
(EPS)

S2 Fig. The genetic configuration evolved with different learning algorithms. Top left: PQL. Top right: QL. Bottom Left: RQL. Bottom right: DRL. All algorithms are able to reproduce the main result of the paper, i.e. the evolution of a generalist configuration.
(EPS)

S3 Fig. Fitness for different combinations of aptitude level and δ for $q > 1$ and plasticity cost $c = 0$. Note that values of $\delta > 0.5$ now maximize the fitness so an evolutionary outcome is possible where a mix of specialists and generalists co-exist. Left: $a_0 = 0.5$, right: $a_0 = 0.6$.
(EPS)

S4 Fig. Fitness for different combinations of aptitude level and δ for $q = 1$ and plasticity cost $c = 0$. Intermediate aptitude levels deliver the same fitness as extreme levels, thus a mixed population will evolve. An intermediate aptitude level of 0.5 is optimal if $\delta = 0.5$, while an extreme aptitude level is optimal for high or low values of δ. Left: $a_0 = 0.5$, right: $a_0 = 0.6$.
(EPS)

S5 Fig. Fitness for different combinations of aptitude level and δ for $0 < q < 1$ and plasticity cost $c = 0$. Intermediate aptitude levels deliver higher fitness than extreme levels, hence specialists have always a lower fitness than generalists. An intermediate a aptitude level is optimal

in any circumstances. Left: $a_0 = 0.5$, right: $a_0 = 0.6$.
(EPS)

S1 File. Supporting information. Document containing supplementary material.
(PDF)

Acknowledgments

The authors would also like to thank Leonard Wossnig and Johannes Thiele for their help in developing the simulation framework. S.B. and L.A. are Joint First Authors of this publication and their names are presented in alphabetical order.

Author Contributions

Conceptualization: Leonel Aguilar, Stefano Bennati, Dirk Helbing.

Data curation: Leonel Aguilar, Stefano Bennati.

Formal analysis: Leonel Aguilar, Stefano Bennati.

Funding acquisition: Dirk Helbing.

Investigation: Leonel Aguilar, Stefano Bennati.

Methodology: Leonel Aguilar, Stefano Bennati, Dirk Helbing.

Project administration: Leonel Aguilar, Stefano Bennati, Dirk Helbing.

Resources: Leonel Aguilar, Stefano Bennati.

Software: Leonel Aguilar, Stefano Bennati.

Supervision: Leonel Aguilar, Dirk Helbing.

Validation: Leonel Aguilar, Stefano Bennati.

Visualization: Leonel Aguilar, Stefano Bennati.

Writing – original draft: Leonel Aguilar, Stefano Bennati, Dirk Helbing.

Writing – review & editing: Leonel Aguilar, Stefano Bennati, Dirk Helbing.

References

1. Baldwin JM. A new factor in evolution. The american naturalist. 1896; 30(354):441–451. https://doi.org/10.1086/276408
2. Ancel LW. Undermining the Baldwin expediting effect: does phenotypic plasticity accelerate evolution? Theoretical population biology. 2000; 58(4):307–319. https://doi.org/10.1006/tpbi.2000.1484 PMID: 11162789
3. West-Eberhard MJ. Phenotypic plasticity and the origins of diversity. Annual review of Ecology and Systematics. 1989; 20(1):249–278. https://doi.org/10.1146/annurev.es.20.110189.001341
4. DeWitt TJ, Scheiner SM. Phenotypic plasticity: functional and conceptual approaches. Oxford University Press; 2004.
5. Via S, Gomulkiewicz R, De Jong G, Scheiner SM, Schlichting CD, Van Tienderen PH. Adaptive phenotypic plasticity: consensus and controversy. Trends in Ecology & Evolution. 1995; 10(5):212–217. https://doi.org/10.1016/S0169-5347(00)89061-8
6. Sterelny K. A review of Evolution and learning: the Baldwin effect reconsidered edited by Bruce Weber and David Depew. Evolution & Development. 2004; 6(4):295–300. https://doi.org/10.1111/j.1525-142X.2004.04035.x
7. DeJager J. Baldwin's Remarkable Effect. Biological Theory. 2016; 11(4):207–219. https://doi.org/10.1007/s13752-016-0250-6

8. Crispo E. The Baldwin effect and genetic assimilation: revisiting two mechanisms of evolutionary change mediated by phenotypic plasticity. Evolution. 2007; 61(11):2469–2479. https://doi.org/10.1111/j.1558-5646.2007.00203.x PMID: 17714500
9. Burkhardt RW. Lamarck, evolution, and the inheritance of acquired characters. Genetics. 2013; 194 (4):793–805. https://doi.org/10.1534/genetics.113.151852 PMID: 23908372
10. Scheiner SM. The Baldwin effect: neglected and misunderstood. The American Naturalist. 2014; 184 (4):ii–iii. https://doi.org/10.1086/677944 PMID: 25354416
11. French R, Messinger A. Genes, phenes and the Baldwin effect: Learning and evolution in a simulated population. In: Artificial Life IV; 1994. p. 277–282.
12. Schull J. Are species intelligent? Behavioral and Brain Sciences. 1990; 13(1):63–75. https://doi.org/10.1017/S0140525X00077785
13. Whitley D, Gordon VS, Mathias K. Lamarckian evolution, the Baldwin effect and function optimization. In: International Conference on Parallel Problem Solving from Nature. Springer; 1994. p. 5–15.
14. Gabriel W, Luttbeg B, Sih A, Tollrian R. Environmental tolerance, heterogeneity, and the evolution of reversible plastic responses. The American Naturalist. 2005; 166(3):339–353. https://doi.org/10.1086/432558 PMID: 16224689
15. Hamblin S, Giraldeau LA. Finding the evolutionarily stable learning rule for frequency-dependent foraging. Animal Behaviour. 2009; 78(6):1343–1350. https://doi.org/10.1016/j.anbehav.2009.09.001
16. Red'ko V, Prokhorov D. Learning and Evolution of Autonomous Adaptive Agents. Advances in Machine Learning I. 2010; p. 491–500.
17. Sznajder B, Sabelis M, Egas M. How adaptive learning affects evolution: reviewing theory on the Baldwin effect. Evolutionary biology. 2012; 39(3):301–310. https://doi.org/10.1007/s11692-011-9155-2 PMID: 22923852
18. Phillips PC. The language of gene interaction. Genetics. 1998; 149(3):1167–1171. PMID: 9649511
19. Panchanathan K, Frankenhuis WE. The evolution of sensitive periods in a model of incremental development. In: Proc. R. Soc. B. vol. 283. The Royal Society; 2016. p. 20152439.
20. Epstein JM. Agent-based computational models and generative social science. Complexity. 1999; 4(5):41–60. https://doi.org/10.1002/(SICI)1099-0526(199905/06)4:5%3C41::AID-CPLX9%3E3.0.CO;2-F
21. Perc M, Gómez-Gardeñes J, Szolnoki A, Floría LM, Moreno Y. Evolutionary dynamics of group interactions on structured populations: a review. Journal of the royal society interface. 2013; 10(80):20120997. https://doi.org/10.1098/rsif.2012.0997
22. Beauchamp G. Learning rules for social foragers: implications for the producer–scrounger game and ideal free distribution theory. Journal of Theoretical Biology. 2000; 207(1):21–35. https://doi.org/10.1006/jtbi.2000.2153 PMID: 11027477
23. Hinton GE, Nowlan SJ. How Learning Can Guide Evolution. Complex systems. 1987; 1(3):495–502.
24. Nolfi S, Floreano D. Learning and Evolution. Autonomous robots. 1999; 7(1):89–113. https://doi.org/10.1023/A:1008973931182
25. Menczer F, Belew RK. Evolving sensors in environments of controlled complexity. In: Artificial life IV. MIT Press; 1994. p. 210–221.
26. Ackley D, Littman M. Interactions between learning and evolution. Artificial life II. 1991; 10:487–509.
27. Bennati S. On the role of collective sensing and evolution in group formation. Swarm Intelligence. 2018;. https://doi.org/10.1007/s11721-018-0156-y
28. Pulliam HR, Dunning JB, Liu J. Population dynamics in complex landscapes: a case study. Ecological Applications. 1992; 2(2):165–177. https://doi.org/10.2307/1941773 PMID: 27759202
29. Beissinger S, Donnay T, Walton R. Experimental analysis of diet specialization in the snail kite: the role of behavioral conservatism. Oecologia. 1994; 100(1):54–65. https://doi.org/10.1007/BF00317130 PMID: 28307027
30. Laverty TM. Costs to foraging bumble bees of switching plant species. Canadian Journal of Zoology. 1994; 72(1):43–47. https://doi.org/10.1139/z94-007
31. Dridi S, Lehmann L. Environmental complexity favors the evolution of learning. Behavioral Ecology. 2015; 27(3):842–850. https://doi.org/10.1093/beheco/arv184
32. Torney CJ, Berdahl A, Couzin ID. Signalling and the evolution of cooperative foraging in dynamic environments. PLoS computational biology. 2011; 7(9). https://doi.org/10.1371/journal.pcbi.1002194 PMID: 21966265
33. Gruau F, Whitley D. Adding learning to the cellular development of neural networks: Evolution and the Baldwin effect. Evolutionary computation. 1993; 1(3):213–233. https://doi.org/10.1162/evco.1993.1.3.213

34. Batali J, Grundy WN. Modeling the evolution of motivation. Evolutionary Computation. 1996; 4(3):235–270. https://doi.org/10.1162/evco.1996.4.3.235

35. Nolfi S, Miglino O, Parisi D. Phenotypic plasticity in evolving neural networks. In: From Perception to Action Conference, 1994. IEEE; 1994. p. 146–157.

36. Watkins CJ, Dayan P. Q-learning. Machine learning. 1992; 8(3-4):279–292. https://doi.org/10.1007/BF00992698

37. Hinton G, Osindero S, Welling M, Teh YW. Unsupervised discovery of nonlinear structure using contrastive backpropagation. Cognitive science. 2006; 30(4):725–731. https://doi.org/10.1207/s15516709cog0000_76 PMID: 21702832

38. Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG, et al. Human-level control through deep reinforcement learning. Nature. 2015; 518(7540):529–533. https://doi.org/10.1038/nature14236 PMID: 25719670

39. Roth AE, Erev I. Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term. Games and economic behavior. 1995; 8(1):164–212. https://doi.org/10.1016/S0899-8256(05)80020-X

40. Erev I, Roth AE. Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. American economic review. 1998; p. 848–881.

41. Cooper DJ, Feltovich N. Selection of Leaming Rules: Theory and Experimental Evidence '. 1997;.

42. Silver D, Schrittwieser J, Simonyan K, Antonoglou I, Huang A, Guez A, et al. Mastering the game of go without human knowledge. Nature. 2017; 550(7676):354. https://doi.org/10.1038/nature24270 PMID: 29052630

43. Giraldeau LA, Caraco T. Social foraging theory. Princeton University Press; 2000.

44. Watson RA, Szathmáry E. How can evolution learn? Trends in ecology & evolution. 2016; 31(2):147–157. https://doi.org/10.1016/j.tree.2015.11.009

45. Frankenhuis WE, Panchanathan K, Belsky J. A mathematical model of the evolution of individual differences in developmental plasticity arising through parental bet-hedging. Developmental science. 2016; 19(2):251–274. https://doi.org/10.1111/desc.12309 PMID: 26010335

46. Van Tienderen PH. Evolution of generalists and specialists in spatially heterogeneous environments. Evolution. 1991; 45(6):1317–1331. https://doi.org/10.1111/j.1558-5646.1991.tb02638.x PMID: 28563821

47. Wilson DS, Yoshimura J. On the coexistence of specialists and generalists. The American Naturalist. 1994; 144(4):692–707. https://doi.org/10.1086/285702

48. DeWitt TJ, Sih A, Wilson DS. Costs and limits of phenotypic plasticity. Trends in ecology & evolution. 1998; 13(2):77–81. https://doi.org/10.1016/S0169-5347(97)01274-3

49. Murren CJ, Auld JR, Callahan H, Ghalambor CK, Handelsman CA, Heskel MA, et al. Constraints on the evolution of phenotypic plasticity: limits and costs of phenotype and plasticity. Heredity. 2015; 115(4):293. https://doi.org/10.1038/hdy.2015.8 PMID: 25690179

50. Padilla DK, Adolph SC. Plastic inducible morphologies are not always adaptive: the importance of time delays in a stochastic environment. Evolutionary Ecology. 1996; 10(1):105–117. https://doi.org/10.1007/BF01239351

51. Kassen R. The experimental evolution of specialists, generalists, and the maintenance of diversity. Journal of evolutionary biology. 2002; 15(2):173–190. https://doi.org/10.1046/j.1420-9101.2002.00377.x

52. Floreano D, Nolfi S. Adaptive behavior in competing co-evolving species. In: 4th European Conference on Artificial Life; 1997. p. 378–387.

53. Wakano JY, Aoki K, Feldman MW. Evolution of social learning: a mathematical analysis. Theoretical population biology. 2004; 66(3):249–258. https://doi.org/10.1016/j.tpb.2004.06.005 PMID: 15465125

54. Lande R. Adaptation to an extraordinary environment by evolution of phenotypic plasticity and genetic assimilation. Journal of evolutionary biology. 2009; 22(7):1435–1446. https://doi.org/10.1111/j.1420-9101.2009.01754.x PMID: 19467134

55. Fordyce JA. The evolutionary consequences of ecological interactions mediated through phenotypic plasticity. Journal of Experimental Biology. 2006; 209(12):2377–2383. https://doi.org/10.1242/jeb.02271 PMID: 16731814

56. Chakra MA, Hilbe C, Traulsen A. Plastic behaviors in hosts promote the emergence of retaliatory parasites. Scientific reports. 2014; 4:4251. https://doi.org/10.1038/srep04251 PMID: 24589512

57. Nolfi S, Parisi D. Learning to adapt to changing environments in evolving neural networks. Adaptive behavior. 1996; 5(1):75–98. https://doi.org/10.1177/105971239600500104

58. Frankenhuis WE, Panchanathan K. Balancing sampling and specialization: An adaptationist model of incremental development. Proceedings of the Royal Society of London B: Biological Sciences. 2011; p. rspb20110055. https://doi.org/10.1098/rspb.2011.0055

59. Ramírez JC, Marshall JA. Can natural selection encode Bayesian priors? Journal of theoretical biology. 2017; 426:57–66. https://doi.org/10.1016/j.jtbi.2017.05.017 PMID: 28536034

60. Chapman BB, Morrell LJ, Krause J. Plasticity in male courtship behaviour as a function of light intensity in guppies. Behavioral Ecology and Sociobiology. 2009; 63(12):1757–1763. https://doi.org/10.1007/s00265-009-0796-4

61. Gottschal JC, de Vries S, Kuenen JG. Competition between the facultatively chemolithotrophic Thiobacillus A2, an obligately chemolithotrophic Thiobacillus and a heterotrophic Spirillum for inorganic and organic substrates. Archives of Microbiology. 1979; 121(3):241–249. https://doi.org/10.1007/BF00425062

62. Gillespie JH. Natural selection for variances in offspring numbers: a new evolutionary principle. The American Naturalist. 1977; 111(981):1010–1014. https://doi.org/10.1086/283230

63. Singh S, Lewis RL, Barto AG, Sorg J. Intrinsically motivated reinforcement learning: An evolutionary perspective. IEEE Transactions on Autonomous Mental Development. 2010; 2(2):70–82. https://doi.org/10.1109/TAMD.2010.2051031

64. Pariser E. The filter bubble: What the Internet is hiding from you. Penguin UK; 2011.

65. Nguyen TT, Hui PM, Harper FM, Terveen L, Konstan JA. Exploring the filter bubble: the effect of using recommender systems on content diversity. In: Proceedings of the 23rd international conference on World wide web. ACM; 2014. p. 677–686.

66. Mäs M, Flache A. Differentiation without distancing. Explaining bi-polarization of opinions without negative influence. PloS one. 2013; 8(11):e74516. https://doi.org/10.1371/journal.pone.0074516 PMID: 24312164

67. Quattrociocchi W, Caldarelli G, Scala A. Opinion dynamics on interacting networks: media competition and social influence. Scientific reports. 2014; 4:4938. https://doi.org/10.1038/srep04938 PMID: 24861995

68. Code Repository: How learning can change the course of evolution;. https://github.com/bennati/baldwin_veering.

69. Code Repository: How learning can change the course of evolution;. https://github.com/leaguilar/baldwin_veering.