Understanding the Role of Naive Learners in Cultural Change

Michael Chimento1,2,3,* and Lucy M. Aplin1,3,4

1. Cognitive and Cultural Ecology Research Group, Max Planck Institute of Animal Behavior, Am Obstberg 1, 78315 Radolfzell, Germany; 2. Centre for the Advanced Study of Collective Behaviour, University of Konstanz, Universitätstraße 10, 78464 Konstanz, Germany; 3. Department of Evolutionary Biology and Environmental Studies, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland; 4. Division of Ecology and Evolution, Research School of Biology, Australian National University, 46 Sullivan Creek Road, Canberra, Australian Capital Territory 2600, Australia

Submitted September 4, 2023; Accepted January 17, 2024; Electronically published April 15, 2024

Online enhancements: supplemental PDF.

ABSTRACT: A change to a population’s social network is a change to the substrate of cultural transmission, affecting behavioral diversity and adaptive cultural evolution. While features of network structure such as population size and density have been well studied, less is understood about the influence of social processes such as population turnover—or the repeated replacement of individuals by naive individuals. Experimental data have led to the hypothesis that naive learners can drive cultural evolution by better assessing the relative value of behaviors, although this hypothesis has been expressed only verbally. We conducted a formal exploration of this hypothesis using a generative model that concurrently simulated its two key ingredients: social transmission and reinforcement learning. We simulated competition between high- and low-reward behaviors while varying turnover magnitude and tempo. Variation in turnover influenced changes in the distributions of cultural behaviors, irrespective of initial knowledge-state conditions. We found optimal turnover regimes that amplified the production of higher reward behaviors through two key mechanisms: repertoire composition and enhanced valuation by agents that knew both behaviors. These effects depended on network and learning parameters. Our model provides formal theoretical support for, and predictions about, the hypothesis that naive learners can shape cultural change through their enhanced sampling ability. By moving from experimental data to theory, we illustrate an underdiscussed generative process that can lead to changes in cultural behavior, arising from an interaction between social dynamics and learning.

Keywords: social learning, cultural evolution, population dynamics, reinforcement learning, agent-based model.

Introduction

Culture, or shared socially learned traits persisting across generations, is observed in various taxa from birds to cetaceans and primates (Rendell and Whitehead 2001; Whitehead et al. 2004; Aplin 2019; Schuppli and van Schaik 2019; Whiten 2019, 2021). Cultural traits are inherited through social transmission and may compete with each other (Galef 1995; O’Brien and Lyman 2000; Mesoudi et al. 2004), either for expression when traits are not mutually exclusive or for hosts when traits are mutually exclusive. Thus, the distribution and forms of cultural traits can potentially evolve to become more adaptive, in terms of either the trait holder’s fitness or the fecundity of the traits themselves (Alexander 1979; Cavalli-Sforza and Feldman 1981; Boyd and Richerson 1988; Lumsden and Wilson 2005; Whiten et al. 2011; Creanza et al. 2017a; Mesoudi and Thornton 2018; Gruber et al. 2022; Micheletti et al. 2023). Many mechanisms have been proposed to promote adaptive cultural evolution, such as transmission biases, teaching, and metacognitive abilities (Boyd and Richerson 1988; Laland 2004; Tennie et al. 2009; Aoki et al. 2011; Rendell et al. 2011; Dean et al. 2012; Heyes 2016, 2018; Legare 2017; Kendal et al. 2018). Beyond these widely discussed mechanisms, our recent experiment hinted that adaptive cultural evolution can occur as a result of an interaction between the social process of population turnover and learning (Chimento et al. 2021).

Turnover,1 a gradual replacement of individuals in a population, is a common feature of nearly all animal social systems, often caused by fission-fusion movements, dispersal between groups, or generational change (Aureli et al. 2008; Sueur et al. 2011; Silk et al. 2014). Importantly, turnover introduces naive learners into a population, providing potential opportunities for behavior to spread and change. In Chimento et al. (2021), we performed a cultural diffusion experiment that directly compared the evolution of a socially learned foraging behavior in static and turnover groups

* Corresponding author; email: mchimento@ab.mpg.de.

ORCIDs: Chimento, https://orcid.org/0000-0001-5697-1701; Aplin, https://orcid.org/0000-0001-5367-826X.

American Naturalist, volume 203, number 6, June 2024. © 2024 The University of Chicago. This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0), which permits non-commercial reuse of the work with attribution. For commercial use, contact journalpermissions@press.uchicago.edu. Published by The University of Chicago Press for The American Society of Naturalists. https://doi.org/10.1086/730110

1. This use of “turnover” differs from the introduction of novel traits (Bentley et al. 2007; Acerbi and Bentley 2014).
of great tits (*Parus major*). When naive individuals were introduced into turnover groups, they preferentially adopted a more efficient foraging behavior over a less efficient established tradition in seven of nine replicates. Static groups innovated but did not adopt the more efficient behavior in eight of nine replicates. Our results supported the hypothesis that new learners might drive cultural evolution by better sampling the available behavior space without interference from prior experience. This hypothesis was originally articulated over three decades ago by Warner (1988, 1990). Warner studied the remarkably consistent spawning site preferences of blue-headed wrasse (*Thalassoma bifasciatum*), which he hypothesized were culturally inherited. He replaced entire subpopulations of fish with naive individuals, which then chose more suitable spawning sites than the previous generation. These new sites were chosen again after another replacement event, suggesting that the original populations’ preferences were the suboptimal result of cultural inertia. This specific interaction between turnover and learning has only been verbally stated, not theoretically formalized or explored, motivating us to do so using a mechanistic computational model.

This hypothesis hinges on the assumption that two types of learning occur simultaneously: (1) the social transmission of repertoire, or acquiring a set of available behavioral options (Hoppitt and Laland 2013), and (2) the individual reinforcement learning that takes place during maintenance of the repertoire (also referred to as guided variation; Boyd and Richerson 1988; Galef 1995). Reinforcement learning results in changing preferences over time as individuals learn the value of behavioral options, which may lead to some behaviors being preferred over others. The importance of both types of learning was highlighted by Rendell et al.’s (2011) social learning strategies tournament, where agents could add behaviors to repertoire through observation and used various mechanisms to decide which behavior to exploit. In Chimento et al. (2022), we presented the case for considering both types of learning when analyzing cultural diffusions, since it resulted in novel patterns of diffusion compared with fundamental diffusion models (Bass 1969; Mahajan and Peterson 1985). With the present study, we aim to demonstrate why linking social transmission and reinforcement learning should also matter for cultural evolution.

Social structures and processes have been studied before in theoretical models of evolution. For example, the effects of gene flow via migration on genetic diversity have been explored in classic population genetics models such as Wright’s island model (Wright 1931). Cultural evolution is likewise influenced by social factors (Nunn et al. 2009; Powell et al. 2009; Lehmann et al. 2011; Whitehead and Lusseau 2012; Cantor et al. 2021) and can even feed back into network structure (Ghirlanda and Enquist 2007; Cantor et al. 2015; Smolla and Akçay 2019). Prior studies highlighted the importance of social factors for cultural diversity and adaptive evolution, such as population size (Henrich 2004; Kobayashi and Aoki 2012; Raviv et al. 2019; Cantor et al. 2021), population density (Powell et al. 2009), network connectivity (Derex and Boyd 2016; Migliano et al. 2020; Cantor et al. 2021), age structure (Acerbi et al. 2012; Deffner and McElreath 2022), population growth (Fogarty et al. 2017), and migrations and population contact (Powell et al. 2009; Creanza et al. 2017b; Somveille et al. 2018). These models of immigration and population contact, while relevant for the study of turnover, focused on transmission and neglected reinforcement learning. A notable exception is the model of guided variation by Boyd and Richerson (1988), which showed that some amount of individual learning hastened the rate of evolution toward an equilibrium point. However, this model did not include horizontal transmission and overlapping generations, which are key ingredients of our hypothesis.

Models of cultural evolution intrinsically include generational turnover, as cultural evolution typically requires new generations of learners. However, turnover has been a fundamental assumption rather than a primary research focus. Some of the most influential models include turnover without overlapping generations; instead, all individuals are replaced simultaneously (Cavalli-Sforza and Feldman 1981; Boyd and Richerson 1988). A more realistic model of turnover with overlapping generations is the Moran process (Moran 1958). Moran models have been used to study the effect of different learning strategies and transmission biases on cultural evolution (Whitehead 2007; Aoki et al. 2011) and cultural extirpation (Kobayashi et al. 2021). However, these studies did not vary the rate or magnitude of turnover. Similarly, cultural evolution experiments have included turnover as part of their design without investigating turnover parameters (Mesoudi and Whiten 2008; Whiten and Mesoudi 2008).

Models have often implemented vertical inheritance of traits, passed directly from parents to offspring at birth. Aoki et al. (2011) tested oblique transmission of traits, which could be inherited from any connected agent. However, transmission was not open ended, and once acquiring a trait, agents could not replace it within their lifetime. In contrast, Strimling et al. (2009) allowed for open-ended transmission but assumed a mutually exclusive trait where agents knew only one behavior at a time. Limiting repertoire to a single, immutable behavior excludes the influence of reinforcement learning. A model without open-ended transmission and competition of multiple behaviors within an individual might not fully capture cultural dynamics in either static populations or interstitial periods between turnover events. With the present study, we aim to compare changes in trait frequencies between static populations and populations with varying levels of turnover. Thus, we propose a modified model that allows for a mechanism for
reinforcement learning (i.e., agents learn the value of behaviors through experience), variation in turnover (i.e., the tempo and magnitude of replacement can vary), open-ended transmission (i.e., agents are able to acquire new behaviors throughout their life), and a repertoire of more than one behavior.

Given the ubiquity of turnover in social species and its central role in cultural evolution, it is crucial to further our understanding of its influence in socially learning species. We created a simple model of turnover using unidirectional migration similar to an island-continent model, where immigrants migrated from a source of naive individuals into a focal population. Within this dynamic network of social learners, we simulated competition between a low-reward and high-reward cultural behavior. Behaviors were not mutually exclusive: agents could hold repertoires with none, one, or both behaviors. The acquisition of behavior was conditioned on its usage, whereby more frequently used behaviors spread faster. During each simulation, we recorded key metrics, such as production and repertoire frequencies, rewards received, and agents’ valuation of each behavior. In doing so, we illuminated the mechanisms that link turnover and reinforcement learning to cultural evolution.

**Methods**

Our simulation’s design was based on our prior experiment, where a high-reward foraging behavior was allowed to compete against a low-reward established tradition in 18 experimental populations, split into static and turnover conditions. We replicated and extended our experiment in silico by simulating networked populations of learning agents that could acquire and use a low-reward behavior \( a \) and a novel, high-reward behavior \( b \). We varied the intensity of turnover by changing (1) the magnitude, or what percentage of the population was replaced, and (2) the tempo, or how frequently replacement events occurred. We further varied network size, density, the transmissibility of behaviors, and the rules that agents used to make behavioral decisions, with the expectation that the effect of turnover might depend on these variables.

**Agent Description**

Our model simulated a population of agents that were characterized by a behavioral repertoire \( Z \). At each time step, agents had the opportunity to socially learn unknown behaviors from their associates and add them to their repertoire. Agents could be in three states: completely naive \( (Z = \emptyset) \), monobehavioral \( (Z = \{a\}) \) or \( Z = \{b\} \), or dibehavioral \( (Z = \{a, b\}) \). Behaviors were added to an agent’s repertoire by sampling behaviors produced by their neighbors. This dynamic ensured that acquisition was conditioned on production, as in our prior computational model of cultural diffusion (Chimento et al. 2022). Our model was agnostic to the mechanism responsible for social transmission and could represent anything from enhancement to imitation.

At each time step, knowledgeable agents also selected one behavior from their repertoire to perform. Agents did not choose behavior randomly but instead learned from experience via reinforcement learning. This was implemented as a simple model of reinforcement learning using a linear operator dynamic (Bush and Mosteller 1951; Sutton and Barto 2018). Thus, agents chose behaviors from their repertoire on the basis of their history of previously received rewards. Over time, agents’ expected values about behaviors updated with experience, influencing the probabilities of choosing behaviors from their repertoire. Since this was a model of intragenerational turnover, received rewards did not influence agents’ fitness. If an agent was naive \( (Z = \emptyset) \), no behavior was performed. If an agent was monobehavioral, the behavior in repertoire \( Z \) was performed. Monobehavioral agents still updated their value function despite having only one behavior to choose from. If an agent was dibehavioral, behavior \( a \) or \( b \) was performed on the basis of their expected values. The equations of this model are described below in “Submodel of Behavioral Production.”

**Network Description**

As shown in figure 1A, the population was organized in a social network, characterized by a size \( n \) and connectivity \( k \) that described the number of associates each agent had. We used random regular networks, where agents were randomly connected to \( k \) other agents. We chose random regular architecture because (1) it kept the effective population size constant across different network densities (Deffner and McElreath 2022) and (2) it meant that each agent had the same number of associates. This was important to control for because variation in the number of associates could directly influence the probabilities calculated in the submodel of acquisition (eq. [2]).

The network underwent a series of turnover events, where agents were periodically replaced by new individuals. We varied the intensity of turnover either by leaving populations as static or by replacing a certain number of agents (turnover magnitude, \( \tau_m \)) at a set interval (turnover tempo, \( \tau_t \)). Each combination of these two parameters defined a turnover regime. Naïve agents replaced older agents with a probability proportional to their time spent in the population and inherited the network connections of their predecessors.

**Initial Conditions**

We tested two different initialization scenarios.

*Initialization 1.* A case similar to our experiment, where all agents held \( Z = \{a\} \) and one random seed
agent held $Z = \{a, b\}$. It resembled the moment when an innovation occurred and would begin to compete against an established tradition. This matched the state of our prior experiment after $a$ had already spread, at the moment when $b$ was first innovated. Rather than including a stochastic innovation rate, we initialized with a seed agent to remove stochastic variance that would have resulted from differences in time taken to innovate $b$.

Initialization 2. A more general case where simulations were initialized with one random seed agent with
Z = \{a\} and another with Z = \{b\}, and all other agents held Z = \emptyset.

Submodel of Acquisition
At each time step, agents could potentially socially learn behaviors a or b from their neighbors, as defined by the social network. We did not allow for asocial innovation. If an agent did not have both behaviors in their repertoire (i.e., Z = \emptyset, Z = \{a\}, or Z = \{b\}), it would observe one random behavior produced by its associates. After observation, the observed behaviors would then be added to its repertoire with a probability defined by parameter \( p_n \), the probability of social transmission, or transmissibility for short. This could be interpreted as representing the difficulty of socially learning a behavior, or the species’ propensity to socially learn.

This meant that the probability of socially learning behavior \( x \) at any time step was simply the product of transmissibility \( (p) \) and the probability of observing \( x \) from the productions of others:

\[
P(\text{agent } i \text{ acquires } x) = p \times P(\text{agent } i \text{ observes } x). \tag{1}
\]

The probability of observation was the proportion of behavior \( x \) that was produced by associates:

\[
P(\text{agent } i \text{ observes } x) = \frac{\sum_{j=1}^{n} A_{ij} U_{xj}}{\sum_{j=1}^{n} A_{ij}}. \tag{2}
\]

Variable \( A_{ij} \) was a binary variable that indicates association between focal agent \( i \) and agent \( j \) \((a_{ij} = 1 \text{ if connected, else } 0)\). The usage of behavior \( x \) by agent \( j \) at time \( t \) was given by variable \( U_{xj} \). If agent \( j \) produced behavior \( a \) at time step \( t \), \( U_{xa} = 1 \) and \( U_{xb} = 0 \). If an associate was completely naive \((Z = \emptyset)\), it still contributed to the denominator of this equation. Thus, acquisition was condition on production: the probability of learning a new behavior depended on who an agent was neighbors with and what those neighbors produced. An agent was more likely to learn behavior \( x \) if many of its associates were actively choosing behavior \( x \). This linked acquisition to production, detailed in the next section.

Submodel of Behavioral Production
Our submodel of behavioral production was a reinforcement learning model with two parameters: learning rate \( \rho \) and risk appetite \( \alpha \). At each time step, agents produced one behavior \( x \) from their repertoire using the submodel of production.

While agent \( i \) knew only one behavior \( x \), the probability of producing it was always \( P(\text{agent } i \text{ uses } x) = 1 \). When agents held both behaviors in repertoire, the current expected values \( Q_{i}(a) \) and \( Q_{i}(b) \) were transformed into the probabilities of producing either behavior using a softmax function (Gibbs 1902; Sutton and Barto 2018):

\[
P(\text{agent } i \text{ uses } x) = \frac{\exp(Q_{i}(x)/\alpha)}{\sum_{x \in \{a,b\}} \exp(Q_{i}(x)/\alpha)}. \tag{3}
\]

This ensured that \( P(\text{agent } i \text{ uses } a) + P(\text{agent } i \text{ uses } b) = 1 \). The risk appetite of the agents was determined by \( \alpha \). If \( \alpha < 1 \), agents were risk averse and would prefer to produce the behavior for which they held the highest expected value. If \( \alpha > 1 \), agents were risk tolerant and would engage in exploratory sampling of behavior, insensitive to expected values. When \( \alpha = 1 \), agents were linearly sensitive to differences in expected values.

At the first time step of a simulation, knowledgeable agents held no expectations about the value of any known behavior, and \( Q_{i}(x) = 0 \). If an agent knew \( a \) and \( b \), this gave it an equal chance of producing either behavior. Upon acquisition of behavior \( x \) at time \( t \), \( Q_{i}(x) = 0 \). This conservatively assumed that agents did not attend to rewards before acquisition and held no prior expectations about the reward.

When agents produced a behavior, they received a deterministic reward, defined as

\[
r_a = 1 \quad \text{and} \quad r_b = 2. \tag{4}
\]

After producing a behavior, agent \( i \) updated its expected values for the following time step \( Q_{i+1}(x) \) using

\[
Q_{i+1}(x) = \rho(U_{xa} r_a + (1 - U_{xa}) Q_{i}(x))
+ (1 - \rho) Q_{i}(x), \tag{5}
\]

where \( \rho \) was a learning rate parameter that controlled how strongly the received reward \( r \) influenced the expected value \( Q \). The term \( U_{xa} \) as defined above, was equal to 1 if behavior \( x \) was chosen and 0 otherwise. Thus, only the expected value for the chosen behavior was updated, and the other behavior’s expected value remained unchanged. In supplementary text S1 (available in the supplemental PDF), we also considered a scenario where the expected value from the unchosen behavior tended to zero (i.e., expected values were updated without the term \((1 - U_{xb})Q_{i}(x)\), effectively updating the unchosen option with \( r = 0 \) and
causing expected values from unused behaviors to be forgotten).

**Parameter Conditions**
We manipulated turnover magnitude \( \tau_{m} \) from 0% of the population to 100% of the population in increments of 10%. We also varied turnover tempo \( \tau \), from 10 to 100 in increments of 10. We conducted an additional exploration of turnover tempos beyond these values.

We conducted a sensitivity analysis by varying parameters related to network size and connectivity, behavioral transmissibility, and reinforcement learning (summarized in fig. 1B, 1C, 1D). Specifically, we tested three population sizes \( (n \in \{32, 64, 128\}) \), along with three levels of network connectivity by changing the relative proportion of \( k \) against \( n \) \( (k \in \{0.0625n, 0.125n, 0.25n\}) \). For example, a network with \( n = 64 \) and \( k = 0.125n \) yielded \( k = 8 \)—comparable to a network where \( n = 32 \) and \( k = 0.25N \) and where \( n = 128 \) and \( k = 0.0625N \). This allowed us to compare network sizes while controlling for degree and vice versa. Within the submodel of acquisition, we varied the transmissibility of behaviors \( (\rho, \in \{0.05, 0.1, 0.2\}) \). Within the submodel of production, we varied decision-making rules. We tested three different parameterizations of the learning rate parameter \( (\rho \in \{0.01, 0.1, 0.9\}) \) and risk appetite parameter \( (\alpha \in \{0.5, 1.0, 2.0\}) \).

We tested these parameter combinations under several "scenarios," as follows:

**Scenario 1.** Used initialization 1. New agents were obliged to learn one behavior upon entering the population, sampled from the productions of the agent it replaced and that agent's neighbors. This was done by setting \( \rho = 1 \) temporarily upon initialization so that the probability of a new agent acquiring behavior \( x \) was only \( P(\text{individual } i \text{ observes } x) \). As a result, complete cultural extinction was impossible, since agents could not have \( Z = \emptyset \). This closely resembled a Moran model, except that it allowed for open-ended transmission. Importantly, the number of knowledgeable individuals was always equal between static and turnover populations allowing for the fairest comparison between the two.

**Scenario 2.** Used initialization 2. Similar to scenario 1 in that agents were obliged to learn a behavior upon entering the population. If an agent joined the population and had no behavior to sample from, they did not acquire any behavior but still could potentially acquire behaviors at a later time step. Complete cultural extinction was possible, although unlikely and unobserved.

**Scenario 3.** Used initialization 1, although agents were not obliged to learn a behavior upon entering the population and could hold \( Z = \emptyset \). Hence, the probability of new agents immediately acquiring behavior \( x \) was \( P(\text{individual } i \text{ observes } x) \), as in equation (1). Complete cultural extinction was possible.

**Scenario 4.** Used initialization 2. Otherwise similar to scenario 3.

**Scenario 5.** Highly conservative learners. Used initialization 1, and agents could have \( Z = \emptyset \). However, we modified the submodel of production to include an autocorrelation factor, detailed in the appendix. This better reproduced the dynamics observed in our original experiment.

We additionally tested the following condition in static populations to compare the effect of information loss from memory and from turnover:

**Scenario 6: Forgetting of value.** Used initialization 2. Agents also updated the expected value of unchosen options with a forgetting rate \( (\rho) \). Discussed in section S1 of the supplemental PDF.

**Analysis**
Our analysis focuses on four key measurements from the model: repertoire frequencies, trait frequencies, expected valuations of behaviors \( (Q) \) in dibehavioral agents, and mean rewards earned per time step, normalized by either population size or the number of knowledgeable agents. For each regime, we then calculated relative reward as

\[
\text{relative reward} = \frac{\text{mean reward (turnover regime)}}{\text{mean reward (static)}} - 1.
\]

When relative reward is less than zero, the regime underperformed static populations with the same parameter settings, and vice versa when more than zero.

To understand how the effect of turnover changed with parameter settings, we conducted a sensitivity analysis in which we varied one parameter at a time while holding all others at the reference setting. Our reference setting was

\[
\begin{align*}
\text{Scenario 1:} & \quad n = 64; \quad k = 0.125n; \quad \rho, = 0.1; \quad \rho = 0.1; \quad \alpha = 1.0. \\
\end{align*}
\]

For each point in parameter space described in "Parameter Conditions," we recorded 500 simulations, each running for 2,000 time steps. At each time step, we recorded these four key measurements. We then compared the results from each turnover regime to those from static populations where all other parameters were the same. This allowed us to assess the relative effect of each turnover regime and how this interacted with other parameters in the model.

**Results**
To understand how population turnover influenced the cultural states of populations, we compared key metrics
of behavioral frequencies, repertoire frequencies, average rewards, and agents’ expected values of behaviors between turnover and static populations under different scenarios. Throughout the main text, we mostly visualize results from scenario 1, which was initialized to represent the moment a novel, high-reward innovation begins to compete against an established tradition. We found that initial conditions generally did not impact the results, and we found largely similar dynamics among scenarios 1–4. We withhold discussion of scenario 5 until the final section. This was the most specific parameterization of the model yet also yielded dynamics most similar to our original experiment. Detailed summary statistics for the five main scenarios at reference level can be found in tables S1–S5 (available online).

**Dynamics in Static Populations**

Scenarios 1 and 3 used the same initial conditions, resulting in identical static conditions. Behavior $b$ took roughly 300 to 700 time steps to spread through the population, and all agents became dibehavioral. Trait production frequencies then stabilized, with agents preferring behavior $b$ around 73% of the time (fig. 2C). The relative value of $b$ to $a$ stabilized at 1, matching the true reward difference (fig. 2E), and agents received a mean reward of $1.66 \pm 0.32$ every time step. In scenarios 2 and 4, both behaviors spread from seed agents. Initially, this led to a mixture of monobehavioral $a$, monobehavioral $b$, and dibehavioral repertoires (fig. 2A). However, the makeup of populations soon became completely dibehavioral in roughly 250 to 350 time steps, faster than in scenarios 1 and 3. Behavior $a$ was not an established tradition but instead spread concurrently with $b$. This removed the relative advantage behavior $a$ had in terms of $Q$ value. Both behaviors were more likely to be produced and thus spread, at least until agents learned the full values of $a$ and $b$.

As expected from our prior model of cultural diffusion in populations of reinforcement learning agents, the learning rate and risk appetite parameters affected these metrics in all scenarios (Chimento et al. 2022). Populations that were more risk averse experienced slower diffusion but strongly preferred the high-reward behavior. A reduced learning rate also saw a slower diffusion and stabilization of expected values and production frequencies. We expected that these parameters might alter turnover effects, because more frequently produced behaviors would be more likely to be transmitted and agents would be time limited in how long they could learn value.

**Optimal, Neutral, and Suboptimal Regimes**

Before addressing the specific dynamics of turnover populations, we note that turnover could either raise or lower relative rewards compared with static populations, contingent on the regime, as depicted in figure 3A and figure S1 for all scenarios. When relative reward was normalized by total population, turnover offered no advantage in scenarios 3 and 4 because agents with empty repertoires received no rewards. However, advantageous regimes were observed when only considering knowledgeable agents and the gradient was similar to that in scenarios 1 and 2 (fig. S2). This raised the question of how cultural populations should be evaluated in light of naive individuals, which we discuss in section S1 of the supplemental PDF.

Slow-tempo, low-magnitude regimes resembled static populations. Relative reward rose with both increasing magnitude and tempo, up to a certain threshold: rapid temps outperformed static at smaller magnitudes, and larger magnitudes excelled at slower tempos. Beyond this threshold, increased turnover intensity led to diminished relative rewards. In the most intense turnover regimes of scenarios 3, 4, and 5, agent rewards plummeted to zero because of behavioral extirpation (fig. S1B, S1C, dark purple). With slower temps, populations could sustain higher magnitudes without negative consequences. Broadening our tempo range up to 2,048 time-step intervals revealed that as tempo dropped, the influence of turnover magnitude diminished, with behavioral frequencies nearing static populations (fig. S3).

**Dynamics in Turnover Populations**

Figure 3A illustrates the mean rewards per agent relative to static populations, although there was considerable variation underlying these means. Within each regime, evolutionary outcomes were influenced by stochastic simulation variance: some runs led to monobehavioral $b$ fixation, others to $a$, yet others to a stable mix of all three phenotypes. We observed that repertoire recombination and a heightened preference for behavior $b$ in dibehavioral agents boosted relative rewards. However, these gains were offset by cultural extirpation and insufficient time for reinforcement learning, resulting in suboptimal relative rewards. These factors linked turnover to differences in relative reward and remained consistent across scenarios 1–4, except for the lack of complete extirpation to influence scenarios 1 and 2.

We begin with the dynamics of two exemplar optimal regimes from scenario 1. Figure 2 depicts temporal patterns for $\tau_t = 60$ and $\tau_m = 0.8$, yielding a 10% increase in relative rewards. Monobehavioral $b$ and dibehavioral repertoires initially increased as novel behaviors proliferated (fig. 2B). The count of monobehavioral $a$ agents quickly decreased, nearing zero, and behavior $a$ was extirpated in 186 out of 500 simulations. Dibehavioral agents also declined, although more slowly, and represented about
30% of the population. In contrast, the population of monobehavioral $b$ agents increased and stabilized around 60%, oscillating with turnover events, and behavior $b$ became extinct in only 12 simulations. Compared with static populations, behavioral frequencies in turnover populations evolved differently. The proportion of agents using behavior $a$ decreased over time, below the proportion of agents who chose $a$ in static populations.
The proportion of agents who chose behavior $b$ rose above that in static populations. In this case, the dominance of $b$ mainly resulted from the rise of monobehavioral $b$ agents. Notably, dibehavioral agents in turnover populations showed a stronger preference for behavior $b$ over $a$ than those in static populations (fig. 2F). We found very similar dynamics from the same regime in scenario 3. However, because agents could have $Z = \emptyset$, the fraction of agents that could not receive any reward meant that the mean reward per agent was below that of static populations (dynamics visualized in fig. S4). Interestingly, when adjusted for the number of knowledgeable agents, behavior $b$ accounted for a mean proportion of 0.89 ± 0.28 of all behaviors—much higher than in static populations. Despite different initial conditions, we also found similar dynamics under scenarios 2 and 4. One minor exception was that monobehavioral $b$ evolved to represent a larger fraction of repertoires in scenario 4 compared with scenario 3, and populations were less likely to use $a$ as the simulations progressed (fig. S4B, S4C, S4E, S4F).

The above represented a case of higher relative reward driven by a population shift toward monobehavioral $b$. Yet we found cases where higher relative reward was instead driven by dibehavioral agents in the regime $\tau_r = 60$.
and $\tau_m = 0.5$ (fig. 2C), resulting in a 4% increased mean relative reward. Both dibehavioral and monobehavioral $b$ agents maintained stable proportions of the populations. Again, this was similar among scenarios 1–4. Importantly, dibehavioral agents in turnover populations valued behavior $b$ more than their static counterparts, supporting our hypothesis that new learners better sample the behavior space because of inexperience. We found identical dynamics in scenarios 3 and 4 under the same regime. Several factors contributed to this. First, new agents were likely to learn $b$ before $a$, as $b$ was often the dominant behavior at any time step. Consequently, monobehavioral $b$ agents already valued $b$ highly before becoming dibehavioral. Second, dibehavioral agents, even if introduced to $a$ first, were inclined to resample behavior $b$ because of their relative inexperience compared with static dibehavioral individuals.

Importantly, turnover regimes did not consistently lead to increased mean rewards. Lower relative rewards followed when $b$ was extirpated in more simulations than $a$. We visualize some exemplar suboptimal regimes in figure S5. In scenario 1, neither null repertoires nor total cultural extirpation was possible. However, faster tempos or larger magnitudes increased the probability of an early extirpation of $b$, causing undermining results in comparison to static populations in many simulations (fig. S5A, S5B). When both magnitude and tempo were high, dibehavioral individuals valued behavior $b$ less than their counterparts in static populations (fig. S5J) before becoming extinct. When normalized by population size, all turnover regimes in scenarios 3 and 4 resulted in poorer performance than static (see fig. S1C, S1D). However, when only considering knowledgeable agents, we also found cases where regimes performed worse. For instance, scenario 3 (where $\tau_m = 0.9$, $\tau_r = 50$) resulted in only two instances of total extirpation, yet $a$ was roughly 50% more likely to fixate than $b$ (fig. S5C). When $\tau_m = 0.8$ and $\tau_r = 30$ in scenario 4 resulted in about a 20% chance of total extirpation ($N = 106$), although $b$ fixated more frequently ($N = 132$) compared with $a$ ($N = 262$; fig. S5D).

In summary, optimal regimes resulted from favorable (i) repertoire recomposition and (ii) weighting of $b$ over $a$ by dibehavioral agents. Suboptimal regimes resulted when extirpation was complete or when extirpation of $b$ occurred more often than $a$. Tempo and magnitude influenced composition similarly across different initial settings, as shown in figure 4 (see fig. S7 for all scenarios). In scenarios 1–4, increasing tempo or increasing magnitude resulted in a monotonic reduction in the proportion of dibehavioral agents. Changes to the proportion of monobehavioral agents, on the other hand, depended on the scenario. In scenarios 1 and 2, larger magnitudes corresponded with more monobehavioral $b$ agents. Faster tempos increased the proportion of monobehavioral $a$ agents, more so in scenario 1 given that it was an established tradition at the first time step. In scenarios 3 and 4, at extreme turnover magnitude and tempo, monobehavioral agents reached an apex and then declined due to rapid turnover events removing knowledgeable individuals or agents not staying long enough to acquire either behavior.

The Effect of Turnover Depended on Network and Learning Parameters

We expected that performance, in terms of relative reward, might vary on the basis of network parameters of size ($n$) and connectivity ($k$). The full sensitivity analysis is visualized in figure S6. In scenarios 1–4, we found surprisingly little effect of network size when controlling for connectivity on relative reward. However, in scenario 5, smaller networks performed marginally better under weak turnover because $b$ could fixate faster. We found the strongest effect of connectivity in scenarios 3–5. Denser networks generally improved mean rewards under turnover and sparser networks performed worse, especially under strong turnover regimes. Denser networks increased the probability that agents might be connected to an agent that knew the novel behavior, decreasing the probability of its extirpation. Transmissibility ($\rho$) strongly affected all scenarios, with increased values allowing populations to experience faster turnover with less negative influence on relative reward. Reduced transmissibility increased the number of high-intensity regimes with negative outcomes relative to static. In summary, network parameters indirectly modified behavioral extinction and transmission probabilities while transmissibility directly altered it, influencing outcomes. Larger networks reduced the risk of behavior extinction while denser networks and increased behavior transmissibility hastened social transmission, making populations more resistant to aggressive turnover effects.

Parameters related to reinforcement learning, risk appetite ($\alpha$), and learning rate ($\rho$) impacted all scenarios. Risk-tolerant agents benefited less from turnover, with dibehavioral individuals likely to sample behaviors regardless of their value (fig. 3E). Risk aversion slightly elevated the advantage of weak turnover regimes, although the effect of stronger turnover was muted relative to reference. This was because both turnover and static populations tended to choose high-reward behaviors, diminishing relative rewards. This parameterization of the behavioral conservatism model did not fully capture the cultural inertia observed in static populations of our prior experiment, a point revisited in the next section. Last, populations with very slow learning rates ($\rho = 0.01$) derived minimal benefit from turnover, as agents struggled to learn behavior values (fig. 3F). This was confirmed in simulations with no reinforcement learning ($\rho = 0$); agents could never
differentiate the value of behavior, neutralizing turnover’s effect on relative reward.

Reproducing Observed Experimental Dynamics

The inspiration for this model was the verbal hypothesis inspired by our prior experiment with great tits. Great tits in static populations rarely sampled the high-reward alternative, and it became the preferred solution in only one of nine static populations. In contrast, turnover populations switched over to a high-reward alternative faster than static populations after a handful of turnover events (Chimento et al. 2021). Our results so far have contrasted with this pattern, as static populations quickly preferred $b$ and turnover populations did not adopt $b$ before static populations. However, we predicted that a key modification,
detailed in the appendix, would generate data that more closely matched our prior results and verbal hypothesis. We introduced a choice autocorrelation parameter frequently used in reinforcement learning models in humans (Katahira 2015). When this parameter was above zero, agents were more likely to repeat recent choices, slowing static populations’ switch to behavior \( b \). Autocorrelation resulted in strong benefits for weak turnover regimes, even when normalized by population size (fig. 5A). Furthermore, it generated very similar temporal dynamics to our experimental data for both static and turnover populations. Behavior \( b \) slowly spread in static populations, yet agents did not adopt the high-reward behavior (fig. 5B, 5D). Populations benefited strongly from turnover, and monobehavioral \( b \) repertoire evolved to dominate the populations (fig. 5C, 5E). The relative valuation of behaviors by dibehavioral agents in turnover populations rose sharply early in simulations (fig. 5S). This supported our hypothesis as to why turnover should matter more for cultural change in behaviorally conservative species since new learners can be “better” samplers of the behavior space than conservative, behaviorally ossified residents.

Discussion

Our results demonstrate how turnover can interact with reinforcement learning during the maintenance of socially learned behaviors, shaping the evolution of the distributions of repertoires and expressed cultural traits. Our results identify two key mechanisms by which turnover improves relative reward: (1) changing the composition of repertoires to favor monobehavioral agents that knew the high-reward behavior and (2) the exaggerated preference for the high-reward behavior by dibehavioral individuals. However, turnover can be a double-edged sword. When social learning is not obligatory, it decreases the proportion of agents that know any behavior, and higher-intensity turnover regimes lead to frequent behavioral extinction. The transitions between increasing and decreasing relative reward depend on the properties of the network and learners. Finally, as found in our experiment, turnover can greatly benefit populations where choice was autocorrelated, which caused strong cultural inertia in static populations.

By combining transmission and reinforcement learning, our model illustrates the importance of behavioral expression for the propagation and subsequent distribution of traits. Even though the forms of traits could not evolve in our study, our model suggests that expression, the rate of which is determined in part by a value function, can also contribute to changes in the form of traits. Traits that evolve to maximize the value function might have a higher chance of surviving the learning bottleneck to the next generation of learners. In many cases, species should evolve learning mechanisms that place high value on adaptive behaviors (West-Eberhard 2003, p. 339), but valuation may be orthogonal or detrimental to an individual’s fitness (Gruber et al. 2022), resulting in the spread of behaviors that are neutral or maladaptive in terms of individual fitness but adaptive in terms of cultural fitness (Franz and Matthews 2010).

Connections to Other Hypotheses about Naive Learners

Our analysis of expected value in dibehavioral agents provides theoretical support for the hypothesis that introducing naive individuals can drive adaptive cultural evolution simply by being better samplers, overcoming the cultural inertia of knowledgeable individuals (Warner 1988, 1990; Chimento et al. 2021). We believe our results should be generalizable, as our model did not assume any particular transmission mechanism and used a very simple reinforcement learning dynamic that has been widely applied across human and nonhuman animals.

Our study complements existing theories about how new learners can influence cultural evolution. In an experiment on the evolution of homing pigeon routes, Sasaki and Biro (2017) suggested that naive birds introduce innovations, which then are available to selection. Both innovation and sampling mechanisms might operate depending on the behavioral context. Furthermore, our hypothesis complements, rather than contrasts, the hypothesis that informational cognitive biases of naive learners, such as a simplicity bias, shape the evolution of cultural traits. This hypothesis has been tested in the context of language (Kirby et al. 2008; Culbertson and Smolensky 2012; Culbertson and Kirby 2016), music (Verhoeft et al. 2014; Savage et al. 2015; Ravigani et al. 2018), and the evolution of technology (Thompson and Griffiths 2021). Our agents were not biased toward one behavior or the other, yet one could imagine that the expression of any latent informational bias would be amplified in the absence of preferences developed through reward or punishment.

Connections to Prior Models

Our model combined the following key components to formalize and explore the hypothesis that naive learners could drive cultural change: open-ended social transmission, individual reinforcement learning, and dynamic populations that could experience varying levels of turnover. Variations of each component of this model have been implemented in myriad prior studies, inviting comparison.

One well-publicized study that included both social transmission and individual learning was the social learning
Figure 5: Autocorrelation reproduced observed experimental dynamics. Results are visualized for $p_s = 0.2$ (NB: the same dynamics were observed under $p_s = 0.1$, although the system evolved more slowly.) A. Mean relative reward (fill) for each regime. Unlike figure 3, relative reward is normalized by population size, supporting the large benefit turnover yielded when agents were highly conservative. Repertoire and production frequencies in static ($B, D$) and turnover regime ($C, E$) were $t_m = 0.3$ and $t_t = 50$. Similar to our original experiment, static populations innovated but did not adopt $b$. In contrast, turnover populations quickly adopted behavior $b$. 

A. Scenario 5 – with autocorrelation

<table>
<thead>
<tr>
<th>Turnover Tempo</th>
<th>Turnover Magnitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>10%</td>
<td>49% 37% 11% -9% -22%</td>
</tr>
<tr>
<td>20%</td>
<td>46% 46% 39% 28% 11%</td>
</tr>
<tr>
<td>30%</td>
<td>38% 46% 46% 39% 24%</td>
</tr>
<tr>
<td>40%</td>
<td>30% 42% 46% 40% 35%</td>
</tr>
<tr>
<td>50%</td>
<td>22% 37% 43% 40% 34%</td>
</tr>
<tr>
<td>60%</td>
<td>17% 35% 39% 39% 33%</td>
</tr>
<tr>
<td>70%</td>
<td>13% 29% 37% 36% 33%</td>
</tr>
<tr>
<td>80%</td>
<td>11% 25% 33% 34% 30%</td>
</tr>
<tr>
<td>90%</td>
<td>9% 22% 30% 31% 28%</td>
</tr>
<tr>
<td>100%</td>
<td>8% 19% 27% 30% 30%</td>
</tr>
</tbody>
</table>

Relative reward (per agent)

-100 -75 -50 -25 0 25

B. Static ($T_m = 0; T_t = 0$)

C. Turnover ($T_m = 0.3; T_t = 50$)

D. Repertoire

E. Behavior

$Z = \{a\}$

$Z = \{b\}$

$Z = \{a, b\}$

Behavior

$a$

$b$
strategies tournament by Rendell et al. (2011). This was a contest between various strategies that determined when and how agents socially learned, innovated, and exploited behaviors in repertoire. In the context of their model, our reinforcement learning dynamic represents one of many ways of determining which behavior to exploit, although unlike the tournament we disallowed the social observation of payoffs to influence valuation. This was done as a simplification in our model; however, we believe that implementing this would not change our finding that turnover can result in hypervaluation of high-payoff behavior. Furthermore, the tournament’s highest-performing learning strategies demonstrated that forgetting is a key requirement for adaptive learning, also supported by a recent model by Ammar et al. (2023). Turnover in our model also represents a mechanism for forgetting cultural memory at a population level. In section S1 of the supplemental PDF, we showed that forgetting of value could increase the usage of higher reward behavior, even without repertoire loss. We emphasize that forgetting alone was not a sufficient model of turnover, since it could not capture changes to repertoire composition. We speculate that any production dynamic that allowed for forgetting of less frequently used repertoire would have caused a to fixate in static populations and exclude the possibility that turnover could outperform it.

A recent model of cultural evolution by André and Baumard (2020) also included open-ended transmission as well as life history decisions, or investments of agents in cultural knowledge. They found that culture could accumulate if it generated improvement in the efficiency of making future investments. Investments in efficient behavior increased the time budget of their agents to innovate or learn other behaviors. In our model, there was no time budget, as agents simply performed one behavior per time step. However, if there were a time budget, along with a mechanism for innovation or alteration of behavioral form, one could generate a similar result from our model. If populations experienced a beneficial regime of turnover, this would result in faster cultural accumulation, as we demonstrated that moderate turnover could result in increased usage of more adaptive behavior, which would free up agents’ time budget.

Another similar study is found in Boyd and Richerson’s model of guided variation, which tested how combining social transmission and individual learning would affect cultural change over intergenerational timescales (Boyd and Richerson 1988, pp. 94–98). Their model illustrated how individual learning hastened the evolution of initial phenotypes in a direction best fit to the local environment, relative to when there was no individual learning (Boyd and Richerson 1988, p. 98). Our model similarly demonstrated that reinforcement learning could change behavioral distributions to become more adaptive, although we extended their model as theirs could not account for intragenerational changes to population structure. Our focus on such intragenerational dynamics—the introduction of naive learners via turnover—highlights the important interaction between population dynamics and reinforcement learning: namely, the introduction of naive learners can potentially amplify the relative proportion of high-reward behaviors beyond the equilibrium point reached in static populations. This could result in an even faster shift in behavioral distributions than predicted by their model.

While we implemented a dynamic population, we did not consider migration between multiple populations in the current model, such as Somveille et al. (2018). In that case, the risk of cultural extinction might be mitigated by movement between populations. Behavioral information would be distributed over larger portions of the metapopulation, buffering against acute cultural extinction. We also did not consider reward-based replacement. In intragenerational turnover, dissatisfied individuals would likely have a higher probability of replacement. Without a mechanism for rewiring, new agents inherit network connections (as in our model) and would likely also experience dissatisfaction. We speculate that feedback between reward and network structure may result in scale-free networks and may differ from network effects related to selection for generalists and specialists found by Smolla and Akçay (2019).

**Turnover at Evolutionary Timescales**

While this study is concerned with the cultural dynamics of intragenerational turnover, we can speculate about evolutionary timescales. Models by Whitehead (2007) and Whitehead and Richerson (2009) show that the capacity for horizontal cultural transmission was adaptive in environments that varied over larger timescales relative to generation times (i.e., red noise environments). Social systems that result in turnover could therefore be most beneficial for socially learning species in the relatively stable periods in between large environmental changes. Interestingly, this contrasts with Sueur et al.’s (2011) prediction that fission-fusion dynamics evolve under “white noise” variability. This discrepancy suggests that the evolution of fission-fusion social behavior should take into account the benefits of social learning.

The loss of dibehavioral individuals under strong turnover regimes could lead to traps at longer timescales, especially if behavior is costly to acquire (either socially or asocially) or if there is no long-term retention of information (e.g., via vertical transmission). In such a case, populations could be composed of monobehavioral individuals that are best fit to the previous environmental state. Despite
this, drastic turnover can still improve fitness when the reinvention of behavior is cheap and obligatory, as illustrated by the example of naive blue-headed wrasse choosing more optimal spawning sites, overcoming cultural inertia (Warner 1988, 1990). Testing variable environments was beyond the scope of this article but would be a fruitful future direction.

Stable social groups and longer memories have been hypothesized to be more adaptive in variable environments, as they allow for the accumulation of rare but important knowledge, such as in matriarch elephants (McComb et al. 2001), and give ample opportunity for the social transmission of complex behaviors (Gruber et al. 2022). However, such species are also often characterized by behavioral conservatism (Whitehead et al. 2004; Marshall-Pescini and Whiten 2008; Hrubesch et al. 2009; Hopper et al. 2011; Davis et al. 2016) and may make maladaptive, conserved decisions, especially when innovation is costly (Keith and Bull 2017; Whitehead and Richerson 2009). This conservatism may be detrimental to the spread of novel behavior, and even if some individuals possess a large repertoire of behavioral knowledge, the realized usage of such knowledge might be constrained by conservative decision-making rules (Chimento et al. 2022). We found that when conservatism was implemented as a risk appetite bias, turnover did not benefit risk averse populations as much as risk-neutral populations. Yet when we implemented it as autocorrelation, performance drastically improved. This suggests that there is nuance to the hypothesis that population turnover is a key mechanism of cultural evolution in behaviorally conservative species in that it might depend on the type of conservatism.

Conclusion

In summary, our study has provided a more thorough understanding of how turnover influenced the internal states of individuals, which then fed back into population-level measures of cultural distributions. Intragenerational turnover is commonly found in many species via fission-fusion dynamics and dispersal. Our results suggest that this is an important mechanism of cultural evolution. Turnover could both enhance and diminish mean rewards compared with static populations depending on the specific regime, with clear thresholds that depended on network and learning parameters. We found these effects with the minimal assumptions that (1) agents could learn from reward and (2) transmission was conditioned on production. This distinction between transmission and production, first introduced in Chimento et al. (2022), offers a useful framework to explore the consequences of transmission and reinforcement learning for culture.

Acknowledgments

We thank Wataru Toyokawa, Iain Couzin, and Rob Boyd for their comments on the manuscript and Erol Akçay, Laurent Lehmann, and two anonymous reviewers for their insightful and helpful comments. M.C. received funding from the IMPRS-QBEE and the Centre for the Advanced Study of Collective Behaviour, funded by the Deutsche Forschungsgemeinschaft (DFG) under Germany’s Excellence Strategy (EXC 2117-422037984). L.M.A. was funded by a Max Planck Research Group Leader Fellowship from the Max Planck Society. This work was additionally supported by the Swiss State Secretariat for Education, Research, and Innovation (SERI) under contract MB22.00056.

Statement of Authorship

Conceptualization: M.C., L.M.A.; methodology: M.C.; software: M.C.; investigation: M.C.; resources: M.C.; writing—original draft: M.C.; writing—review and editing: M.C., L.M.A.; visualization: M.C.; supervision: L.M.A.; funding acquisition: L.M.A.

Data and Code Availability

Code and data for statistical analyses and main text figures, as well as code to replicate the agent-based model, are available for download on Edmond (https://doi.org/10.17617/3.RLRJZC; Chimento 2023).

APPENDIX

Additional Parameterizations

Modification required to simulate experimental dynamics. In the main text, we simplified the model as much as possible. However, the flexibility of our model allows for limitless parameterizations of either submodel. We detail the change made to better capture the dynamics found in our prior experiment with great tits. We introduced a choice autocorrelation factor by modifying equation (3) to the following:

\[
P(\text{agent } i \text{ uses } x) = \frac{\exp(Q_i(x) + \gamma C_i(x))}{\sum_{i\in(ab)} \exp(Q_i(x) + \gamma C_i(x))},
\]  

(A1)

where

\[
C_i(x) = \sum_{i=m} C_m(x).
\]  

(A2)

Here, \(C_m(x)\) recorded how frequently the focal individual produced behavior \(x\) within sliding window \(m\). Parameter
\( \gamma \) controlled the tendency of avoiding or repeating recently chosen behaviors, depending on its sign and value. We set \( \gamma = .05 \). We set \( m = 100 \), and note that larger \( m \) values would result in a stronger effect of \( \gamma \) at a given value.

Literature Cited


The oldest species of Protolabis, *P. transmontanus Cope* [figured], was obtained from the Ticholeptus beds, which overlie the John Day beds in Central Oregon. Its skull and a few bones only are known, but the former displays very complete dentition. Its size is about that of the Virginia deer.” From "The Phylogeny of the Camelidae" by E. D. Cope (*The American Naturalist*, 1886, 20:611–624).