Computational principles underlying the evolution of cultural learning mechanisms

Xavier Roberts-Gaal (xavierrobertsgaal@g.harvard.edu) and
Fiery Cushman (cushman@fas.harvard.edu)
Department of Psychology, Harvard University
William James Hall, 33 Kirkland Street, Cambridge MA

Abstract

Cumulative culture requires efficient learning mechanisms that can withstand environmental change across generations. We unify two competing theories of the learning mechanism supporting cumulative culture in a common computational framework, distinguishing model-based from model-free social learning. We describe and analyze evolutionary models that explain when and why model-based and model-free social learning are each optimal, and in particular how *environmental volatility* determines which strategy succeeds. Strikingly, we find that model-based social learning can succeed even in high-volatility environments. These results yield novel predictions concerning cultural variation in social learning mechanisms.

Keywords: social learning; model-based learning; evolutionary models; cultural learning; imitation; emulation

Introduction

Cumulative culture, the accretion over generations of improvements in practices, artifacts, and technologies, is a distinctive ingredient in human success (Henrich, 2017). Through culture, people can rapidly learn useful behaviors so complex that no individual could invent them from scratch, and then transmit them accurately to future generations (Tomasello, Kruger, & Ratner, 1993).

Currently two models of cultural learning predominate (Heyes & Moore, 2021). One view posits cultural learning is pure imitation: copying others' behaviors precisely, like a baker's apprentice who is given an exact recipe (Heyes, 2018; Henrich, 2021). From a computational perspective, imitation could be described as a *model-free* (MF) form of social learning: it represents the value of actions, but doesn't provide a causal model of why they are valuable. This makes it fast and easy to implement, but inflexible when circumstances change.

Another view posits the transmission relies on rich causal models of goals, values, and the world, like an apprentice baker who aims to replicate a supervisor's signature sourdough, who understands that yeast needs warmth and sugar, and who knows that a strong rise makes softer bread. Computationally, this is a *model-based* (MB) strategy (Kleiman-Weiner et al., 2020; Shafto, Goodman, & Griffiths, 2014; Caldwell & Millen, 2009), and it is cognitively demanding but offers flexibility when circumstances change (e.g., sugar is out, but honey is on hand).

Although these have sometimes been regarded as competing theories of social learning, by now there is ample evidence that humans do both (Wu, Vélez, & Cushman, 2022). The next key challenge for the field, then, is to understand the principles that determine which we use, and when.

In this study, we present a new model that explains how *environments* dictate the optimality of a MF or MB social learning strategy. This model predicts populations coordinate on an optimal social learning strategy, which is expressed at an individual level as a preference for MF or MB social learning. We relate this social learning strategy preference to environmental volatility, or instability resulting in a certain probability of change.

Computational Model

Our model aims to determine when social learners should copy behaviors exactly (model-free) and when they should instead extract people's goals and beliefs to devise their own tailored solution (model-based). The core insight of this model is that social learning is fragile when environments change: intuitively, what worked for the last generation won't work in the present, so individual learning is required. However, spending the cognitive effort to fit and use a rich causal model can partially immunize MB social learners against environmental change. Therefore, MB and MF social learning are optimal in different environments, namely environments at different levels of volatility—which can be construed here as the expected change per generation.

A broad tradition of cultural evolutionary modeling (Boyd & Richerson, 1980; Rogers, 1988) describes how environmental volatility can influence choice of learning strategy. Prior models (Giuliano & Nunn, 2021) have identified equilibria between individual learning and social learning in a continuous, deterministic, infinite-population setting called the replicator process (Taylor & Jonker, 1978; Schuster & Sigmund, 1983).

Here, we draw on a related class of models that describe stochastic dynamics in discrete, fixed-size populations (Nowak, Sasaki, Taylor, & Fudenberg, 2004) to introduce a general setting that enables us to make novel predictions about the cultural evolution of MB and MF social learning.

Setting The setting consists of a finite population, with each person choosing an action from two possible options $A \in \{a_1, a_2\}$. Our setting introduces novel environmental dynamics, allowing us to distinguish various types of social learning. An action results in one of two observed outcomes $O \in \{o_1, o_2\}$. Which outcome the action results in depends on the state of the world $S \in \{s_1, s_2\}$. If the state is s_1, then action a_1 results in outcome o_1 and a_2 results in o_2.

In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society*.

If the state is s_2, then the action-outcome mappings are instead $\{a_1 \rightarrow o_2,\ a_2 \rightarrow o_1\}$. For simplicity, we model people as choosing only once and we consider the deterministic mapping, where an action always achieves only one outcome.

Outcomes, in turn, generate payoffs: o_1 yields a payoff $R = \beta > 0$ and o_2 yields a negative payoff $-\beta$. Outcomes always yield the same payoff regardless of the world state.

Crucially, the environment can change across generations, disrupting payoffs. In each generation, with probability $\Delta \in [0,1]$ there is a *shock* resulting in a new draw of the state from a uniform distribution. People do not initially know the state of the world, so at the beginning of their lives they are uncertain which action is best.

In our model, we explore the relative success of distinct types of learning strategies as a function of (i) parameters influencing how effective each type of strategy is on its own, and (ii) the prevalence of each type in society.

To do this, we rely on the Moran process, a canonical agent-based model of biological evolution in finite populations (Ewens, 2004). In this process, individuals belong to one of a set of types, each with their own distinct payoff function. Between each generation, an individual dies at random and an individual is chosen to reproduce with probability proportional to their payoff. We assume that descendants do not choose their type; instead, they inherit it. Our model does not assume a specific biological or cultural inheritance mechanism. Thus, the type with a higher payoff will tend to increase in its population proportion, and types with lower payoffs will tend to decrease. We explore the dynamics of the Moran process for various types of social learning strategies under different environmental conditions.

Model 1: Social vs. individual learners Our model allows us to recapitulate, in a stochastic and more general framework, results from Giuliano and Nunn (2021, hereafter GN) identifying when people would engage in social learning at all. We begin by demonstrating that our agent-based simulation of evolutionary dynamics in a finite population recreates their results from the analysis of a replicator process. Then, we proceed to our generalization of this framework to capture the distinction between MF and MB social learning.

In GN there are two types of people: individual and social learners. *Social learners* (SL) copy the *action* of a randomly chosen member of the previous generation, allowing for vertical and oblique transmission. Social learners therefore reproduce, in expectation, the distribution of actions of the previous generation. *Individual learners* (IL) do not copy others, instead they learn for themselves what the best action is through trial-and-error. Although they learn the best action with certainty, they pay a cost of learning $\kappa \in [0,\beta]$.[1]

Since we are interested in learning strategy preferences that evolve over generational timescales, we consider only the steady state population proportion of each learning strategy. To determine these proportions, we can examine the payoffs of each strategy. Since individual learners pay a fixed learning cost to identify the optimal action, their payoff is fixed at:

$$R_{IL} = \beta - \kappa$$

Social learners, in contrast, depend on previous generations for their payoff. If a shock occurs between the learner's and teacher's generations, then the learner's expected payoff is 0 as the learner's action will be optimal with probability 0.5. So, an SL can obtain a positive payoff only if they copy the correct action and if a shock does not disrupt the state of the world anywhere along the chain of teachers and learners.

Table 1: Social learning chains resulting in positive payoff

Learning chain	P(No shocks occurring)
$IL \rightarrow SL$	$1-\Delta$
$IL \rightarrow SL \rightarrow SL$	$(1-\Delta)^2$
$IL \rightarrow SL \rightarrow SL \rightarrow SL$	$(1-\Delta)^3$
$IL \rightarrow [n \times SL] \rightarrow SL$	$(1-\Delta)^n$

Following GN, table 1 lists several possible ways a social learner could obtain positive reward, along with their associated probabilities. Extrapolating from this, the expected payoff of a social learner is β times the sum of a sequence of probabilities (representing all the ways a SL could learn the correct action). In the deterministic case, this resolves to:

$$R_{SL} = \beta \times \sum_{i=1}^{\infty} SL^{i-1}(1-SL)(1-\Delta)^i = \frac{\beta(1-SL)(1-\Delta)}{1-SL(1-\Delta)}$$

where SL denotes the population proportion of social learners (Giuliano & Nunn, 2021). In equilibrium, social and individual learners are both present when their payoffs are equal. Setting $R_{SL} = R_{IL}$ and solving for SL, we get the equilibrium proportion of social learners:

$$SL^* = \begin{cases} \frac{-\Delta\beta+\kappa}{\kappa(1-\Delta)} & \text{for } \Delta \in [0, \frac{\kappa}{\beta}) \\ 0 & \text{for } \Delta \in [\frac{\kappa}{\beta}, 1] \end{cases}$$

This equilibrium is stable: if SL increases above equilibrium, social learners are less likely to copy from an individual learner who has obtained the optimal action with certainty. But if SL decreases below equilibrium, then marginal social learners are more likely to copy from an individual learner without needing to pay a learning cost. Moreover, this equilibrium depends crucially on the learning cost κ and environmental stability $(1-\Delta)$. Intuitively, the equilibrium proportion of social learners increases as environmental stability increases. In more stable environments, the world is less likely to change between the previous and current generation. Social learners are thus more likely to copy an action which remains optimal for the present circumstances, regardless of their population proportion. As stability decreases, the slope of R_{SL} with respect to SL decreases too, yielding a lower equilibrium proportion of social learners.

[1]As GN point out, if $\kappa > \beta$ then there are no individual learners since the cost of learning is prohibitively high.

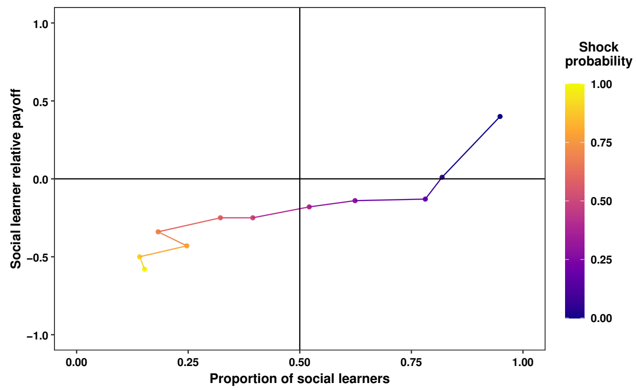

Figure 1: Model 1, relationship between the population proportion and relative payoff of social learners, at varying shock probabilities (Δ). Positive y values indicate SL $>$ IL payoff. Points represent SL averages over 2500 gens from a population of 120 under moderate learning conditions ($\kappa = 0.4$).

We identify the same major result by using a Moran process to model stochastic dynamics. Figure 1 shows the average long-run proportion of social vs. individual learners at varying values of environmental volatility. The long-run proportion of social learners varies directly with the average payoff of social learners relative to individual learners. Moreover, as the shock probability decreases, the expected payoff of social learning increases relative to individual learning and the long-run proportion swings to favor social learners. There are inevitable differences in the dynamics of our model, as the selection mechanism differs and we directly simulate the payoffs of each agent—but overall, the pattern of results holds.

Model 1 allows us to predict that individuals from an environment with *greater cross-generational instability* should have, on average, *reduced preference for social learning*, as compared to individual learning. The consistency of our results in the stochastic setting with prior analytical results in the deterministic setting gives us confidence in extending our approach to consider model-based social learning.

Model 2: Extension to model-based social learning Model 1 assumes that all social learning is model-free social learning; that is, social learners merely copy the action of a randomly-chosen member of the previous generation. Since they copy actions aimlessly, unlike individual learners they cannot detect when the environment changes and therefore cannot adapt. But, there is a wealth of evidence that people also rely on model-based social learning to understand others' goals, environmental dynamics, and then devise their own best approach to novel tasks. MB social learning, unlike MF learning, can partially immunize social learners against environmental change.

Copying another's goal rather than simply their action affords MB social learners the ability to detect when behavior that worked in the past no longer succeeds in the present. Moreover, having a causal model of their environment enables MB social learners to adapt to the change effectively. For example, if you copy the goal of a sourdough boule's texture or flavor, and you learn an understanding of the basic process and ingredients—instead of memorizing precise quantities of flour, water, yeast, mother, temperature, and time—then you can adjust to a higher altitude by baking the bread hotter or longer. If you only followed the bread recipe mechanically, you would not be able to adjust as effectively.

The impact of environmental variability on MB social learning is different from its impact on MF social learning, so we need to develop its payoff function. On top of IL and SL (which we rename MF), we introduce a new type: MB, or model-based social learning. Model-based social learners can copy the *goal* of the person they observe, rather than the action. To achieve this they pay a learning cost λ, corresponding to the difficulty of fitting a model of the environment and the agent's intentions. For example, if they observe $A = a_1, O = o_1$, then they can reliably achieve o_1 even if the state of the world changes. MB social learners have a model of the environmental dynamics and the observed person's goal such that they are immune to environmental shocks occurring in the immediate past.

However, MB social learners are still vulnerable to copying the wrong *goal*. For example, a MB social learner in generation t can copy a MF social learner in generation $(t-1)$ who copied from an individual learner in $(t-2)$. If a shock occurs between $(t-1)$ and t, the MB social learner will still copy the correct outcome and achieve an optimal payoff. But if a shock occurs between $(t-2)$ and $(t-1)$, then the MF social learner has an expected payoff of 0 and so does the MB social learner. Copying someone else's goal guarantees only that you achieve *their* goal, not necessarily the right one.

By construction, MF social learners have a payoff that depends on the relative frequencies of each type in the population as well as the probability of shock. The payoff of MB social learners similarly depends on the population proportions of each type as well as the probability of shock (to a reduced degree) and the model-fitting cost. In the following section we empirically investigate the dynamics of each type.

Having laid out Models 1 and 2, we now analyze Model 2. This enables us to generate testable predictions about the relationship between environmental volatility (Δ) and likelihood of preferring MF, MB, or IL learning strategies on novel tasks.

Results

Inspecting the design of Model 2, we expect the relative prevalence of MF and MB in equilibrium to depend on the relationship between environmental variability (Δ) and model-fitting learning cost (λ). In variable environments, it may prove safer to learn a basic strategy from someone else and tailor it to one's own purposes. However, if it is very hard to fit a causal model of one's environment and to identify the correct goal ($\lambda \gg 0$), then a simpler MF learning strategy works best.

In Model 1, the expected evolutionary equilibria can be characterized analytically, as shown by GN. In contrast, we use an agent-based approach to explore Model 2's evolutionary dynamics, employing a Moran process for selection (Ewens, 2004).

As in Model 1, for Model 2 there exist negative feedback dynamics between social and individual learning. A plethora of individual learners lowers the risk of failure to both social learner types, ensuring they can find the correct action more cheaply than individual learners in their own generation. But, if there are too many social learners, then the risk of copying from someone who has not themselves acquired the correct action increases. In that case, individual learners will have a higher payoff, restoring the equilibrium.

Our key addition in Model 2 is the inclusion of different social learning strategies. As we are interested in understanding when societies would coordinate on these distinct social learning strategies, the evolutionary dynamics of Model 2 are the primary target of our analysis.

We characterize the relative success of individual learning, model-based social learning, and model-free social learning in terms of three parameters: Δ, λ, and κ, each of which can take values in $[0,1]$. Therefore, we exhaustively explore this parameter space to identify interesting regions where we can make novel predictions.

First, we initialize a population of size 120 randomly distributed among learning types.[2] Next, for each value of Δ, λ, and κ we allow the model to cycle 2,500 generations of births and deaths. At each generation, we kill one agent randomly and reproduce one agent in proportion to the softmax of its payoff before allowing agents to choose actions and receive payoffs according to their type. Then, we measure the *concentration* of each type: its average frequency over generations 2,500–5,000, divided by the population size. Finally, we reinitialize and repeat for a new set of parameter values.[3]

We selected this measurement window to allow enough time for populations dynamics to stabilize. To prevent fixation due to drift, we introduce a probability of mutation of 0.05 for each new individual born.

Using this setup, we can understand when environmental volatility, in this case modeled as shock probability, determines the success of different learning strategies. We explore how types' concentrations change as a function of increasing shock probability, Δ, for a given set of model-fitting and individual learning costs—λ and κ, respectively.

We begin by inspecting a case where the cost of individual learning is moderate ($\kappa = 0.4$) and the cost of model-based learning ranges below this value ($0.1 \leq \lambda \leq 0.3$). Figure 2 shows how volatility affects the concentrations of each type.

Several expected results are immediately evident. First, under increased volatility, the concentration of individual learning increases and that of model-free social learning decreases.

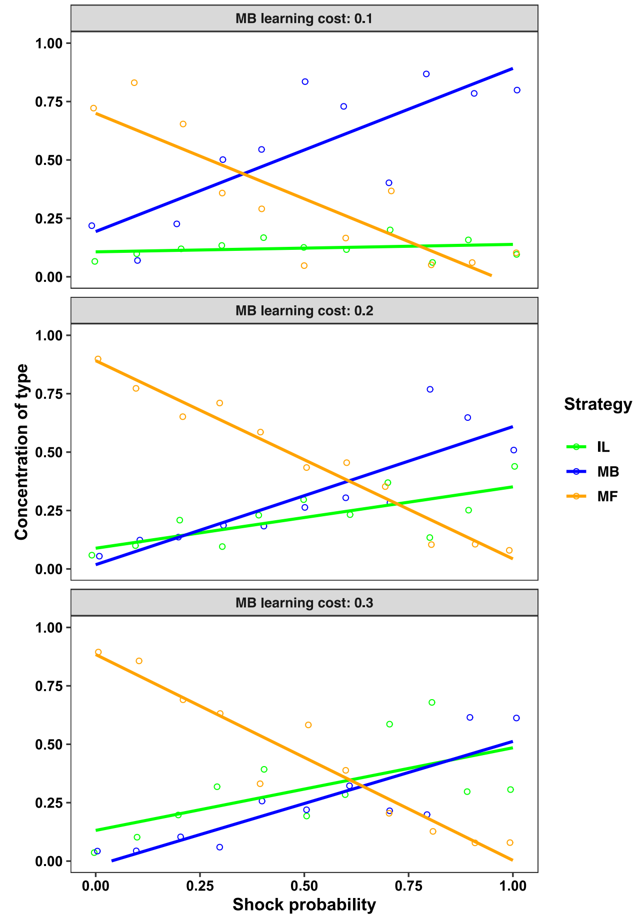

Figure 2: Concentrations of types as shock probability (Δ) increases, for different values of model-based social learning cost (λ). Each collection of three points with the same x-intercept represents the equilibrium of one simulation; lines for each type display linear least-squares trend of concentration with respect to shock probability. All simulations occurred in a moderate learning cost regime with individual learning cost $\kappa = 0.4$.

In the limit, as Δ approaches 1, MF concentration approaches the mutation floor. Second, the relationship between λ and κ influences the concentrations of MB and IL. When κ is very high compared to λ, MB fares better compared to IL.

Moreover, comparing the panels of Figure 2, we find that the effect of volatility (Δ) on concentrations itself depends on the values of λ and κ. That is, as the ratio of MB to individual learning costs $\frac{\lambda}{\kappa}$ decreases (favoring MB social learning), the slope of the MB concentration curve increases relative to the slope of the IL concentration curve, as does its absolute value. Thus, volatile environments favor MB social learning more strongly when social model-fitting costs are lower; they favor individual learning more when it is harder to fit a model.

Counterintuitively, we find that increasing environmental volatility leads to an *increase* in model-based social learning. From Model 1, we might expect that increased volatility

[2]We find similar results using different random seeds.

[3]Reproducible code and data are available at `github.com/xavierrobertsgaal/cogsci-23-computational-culture`

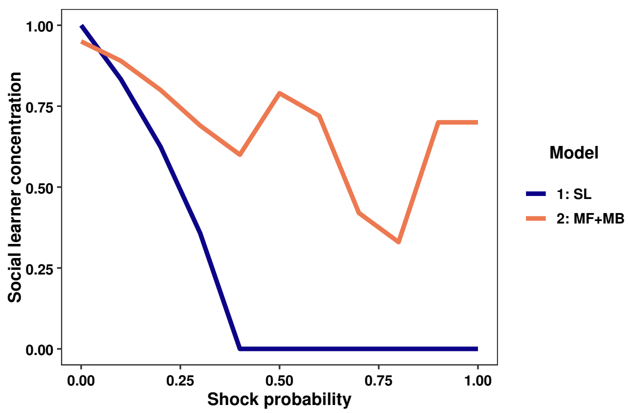

Figure 3: Comparison of social learner concentrations generated by different models as a function of increasing volatility, (Δ). Social learner concentrations for Model 1 use the analytical solution (GN). Social learner concentrations for Model 2 are calculated as the empirical sum of MB and MF, simulated under moderate learning conditions ($\lambda = 0.3$, $\kappa = 0.4$).

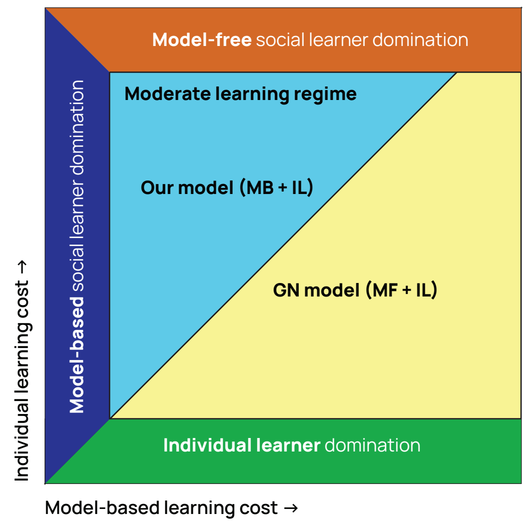

Figure 4: Conceptual framework illustrating distinct learning regimes produced by Model 2 under various social and individual learning conditions. Extreme learning conditions yield populations dominated by one agent type (orange, blue, green). Moderate learning conditions yield populations sensitive to environmental volatility. As volatility increases, agents converge on MB + IL (cyan), or MF + IL (yellow).

leads to monotonically decreasing social learning of any kind, model-based or model-free. After all, both are susceptible to copying social learners who themselves perform the wrong action. Indeed, we find that the proportion of model-free learners dramatically reduces when volatility is large compared to the model-based social learning cost ($\Delta \gg \lambda > 0$). But, since model-based learners are partially immunized from environmental shock, they benefit more when the population shifts to predominantly individual learners. Therefore, the social learning mix shifts to almost exclusively model-based learners, but remains high.

To see this point more clearly, it helps to adopt a coarse-grained division between individual learning and "social learning", comprising both MF and MB approaches. This way, we can compare the fate of social learning when the MB strategy is available (as in our novel model, Model 2) versus when it is not (as in the original GN model, Model 1). As shown in Figure 3, environments in a moderate learning regime ($\lambda = 0.3$, $\kappa = 0.4$) have a high concentration of social learners regardless of environmental volatility. This results from the model-based social learners' distinctive advantage: partial immunity to environmental shock, in exchange for a small, though not trivial, learning cost. In Model 2, as in Model 1, there are many social learners when environmental volatility is low, and these social learners are mostly model-free. High levels of environmental volatility drive these model-free learners to the mutation threshold of near-extinction. However, model-based social learners can survive even very high levels of environmental volatility, since they can reliably identify optimal behavior in a new world state just so long as they have acquired the right goal. For some parameter settings, like those described in Figure 3, model-based social learners can outnumber individual learners on average, even under extremely high volatility.

So far we have investigated key regions of our parameter space in detail. Next, we offer a broad characterization of Model 2 across the full parameter space. We observe the model converges to one of several distinct regimes, depending on the parameter settings. Figure 4 shows these regimes.

Sometimes, only one type will dominate regardless of environmental volatility. In these cases, all other types remain at the mutation floor. When individual learning costs are trivial, for example, the population will consist principally of individual learners, who are not susceptible to shocks and thus never make errors (green region).

When individual and model-based social learning costs are both extreme (e.g., $\kappa = \lambda = 1$), the population converges on model-free social learning regardless of volatility as less error-prone strategies are prohibitively costly (orange region).

When model-based social learning costs are trivially cheap and individual learning is non-trivially expensive, MB learners predominate because they mitigate the risks of shock at a relatively small cost (dark blue region). These extreme learning regimes occur at the boundaries of our parameter space.

However, in less extreme learning regimes, no one type predominates. Instead, both social learners and individual learners coexist, in proportions that are sensitive to environmental volatility. One region of the moderate learning regime encompasses GN's model (Model 1, yellow region in Figure 4). There, the population is almost entirely MF social learners or individual learners, as it has higher model-based than individual learning costs, selecting against MB learners.

Finally, the preceding figures have analyzed a "moderate regime" where model-based learners coexist with model-free and individual learners (cyan region). As shown previously, in this region, increased volatility drops the concentration of model-free learners to the mutation floor. This region is characterized by parameter values ($0 < \lambda < \kappa \ll 1$).

Although these regimes are not sharply delineated due to Model 2's nonlinear dynamics, this conceptual framework reveals a unique finding of our model: that under moderate conditions, social learning can occur in different *forms* (MB or MF) and concentrations sensitive to environmental volatility.

Discussion

When is individual learning versus social learning favored? Prior work identifies environmental volatility as a key variable, with higher levels favoring individual learning. However, by enriching the space of social learning strategies considered, we drive more nuanced predictions about how variations in environmental volatility dictate not only the competition between individual and social learning, but also between alternative forms of social learning. Specifically, we find that when learners are able to copy others' goals while innovating their own means to attain those goals—a variety of "model based" social learning—this expands the range of environments in which social learning can be favored by a sizable portion of the population.

We began by replicating analytical results of a model by Giuliano and Nunn (2021) that contrasts individual learning with non-causal ("model-free") social learning. Our replication shows that its main qualitative findings emerge in an agent-based simulation of evolutionary dynamics as well (Model 1). Examining this model, in contexts characterized by high levels of environmental volatility we find fewer model-free social learners.

However, Model 1 indicated that, all else being equal, environments with more frequent shocks would see a relatively greater proportion of individual learners. Strikingly, we find that sophisticated social learning strategies, such as model-based social learning, can outpace individual learning in some regions of parameter space characterized by high volatility. Specifically, when it is costly to discover which of several goals to pursue, but relatively easy to determine what others' goals are and plan one's own strategy to attain them, social learning can thrive even in the face of environmental change.

Thus, Model 2 allows us to predict that we will (1) observe variation in social learning strategies across human populations that have experienced different levels of historical volatility, and that (2) cultures that have had to adapt to moderate levels of volatility are more likely to adopt a model-based mechanism for social learning, as compared to cultures that have experienced low historical volatility. Future work should test this prediction using data on environmental volatility, such as historical climate variation (Giuliano & Nunn, 2021). Recently, experimental paradigms have been developed to explore the extent to which people engage in goal emulation or action imitation in a social learning task (Charpentier, Iigaya, & O'Doherty, 2020). Fusing these disparate methods may enable us to shed light on systematic variation in preferences for social learning strategy across cultures.

In Model 2, individual learners must be present for cultural learning to succeed. Otherwise, model-based and model-free social learners could copy the wrong goals or actions, respectively. This dynamic is not unlike the Baldwin Effect, whereby learned adaptations to environmental change can acquire a genetic basis (Baldwin, 1896; Simpson, 1953; Heyes, Chater, & Dwyer, 2020). Of course, there are also other ways in which cultural selection, absent individual learning, could nevertheless lead to cultural evolution (e.g., prestige bias, Henrich & Gil-White, 2001; Henrich, 2017).

One important limitation of our current modeling approach is our choice to instantiate environmental shocks by altering the transition structure between actions and observed outcomes when the world state changes. Embedded in this approach is the assumption that shocks change the mappings from actions to outcomes, but not the rewards of outcomes themselves. This allows model-based learners to correct for shocks intervening between the individual they have observed and their own behavior. If, alternatively, shocks changed the unobserved reward values associated with actions, model-based learners of the kind formalized here would have no opportunity to observe or correct for it. Future work should explore the effects of more complicated environmental dynamics—including, for example, weakening or reversing the reward attaching to a specific outcome.

Future work should also seek to test this model's predictions empirically by assessing whether human social learning converges on model-free or model-based formats depending on environmental volatility (Wu et al., 2022).

A second key opportunity for future research is extending our setup to probabilistic mappings between actions and outcomes. Extensive study of model-based social learning has shown that people can invert a Bayesian model to infer a probability distribution over goals and beliefs based on observed behavior (Vélez & Gweon, 2021; Shafto et al., 2014; Baker, Saxe, & Tenenbaum, 2009). Moreover, models in the RL setting are construed as probability distributions over state transitions (Hassabis, Kumaran, Summerfield, & Botvinick, 2017; Dayan & Daw, 2008). Exploring the evolutionary dynamics of cognitive adaptations to probabilistic tasks may therefore illuminate the mechanisms of social inference and cultural learning in an uncertain and changing world.

Conclusion

The present work charts a first step to understanding why, and when, we employ each element of our arsenal of social learning strategies.

Using an evolutionary model, we show that sophisticated social learning strategies—such as model-based inference and Bayesian theories of mind—may play a crucial role in insulating social learners from environmental volatility, thereby enabling cultural advances in one generation to crystallize and form the foundation for further progress in the next.

Acknowledgments

We thank Adam Bear for discussion of the replicator system and of Model 2. We also thank Fred Callaway, Arunima Sarin, Arthur Le Pargneux, Natalia Vélez, and other members of the Moral Psychology Research Lab for their suggestions, as well as the reviewers for their feedback. This work was supported by award number N00014-22-1-2205 from the Office of Naval Research (to FC).

References

Baker, C. L., Saxe, R., & Tenenbaum, J. B. (2009). Action understanding as inverse planning. *Cognition*, *113*(3), 329–349. doi: 10.1016/j.cognition.2009.07.005

Baldwin, J. M. (1896). A New Factor in Evolution. *The American Naturalist*, *30*(354), 441–451.

Boyd, R., & Richerson, P. J. (1980). Sociobiology, culture and economic theory. *Journal of Economic Behavior & Organization*, *1*(2), 97–121. doi: 10.1016/0167-2681(80)90005-0

Caldwell, C. A., & Millen, A. E. (2009). Social Learning Mechanisms and Cumulative Cultural Evolution: Is Imitation Necessary? *Psychological Science*, *20*(12), 1478–1483. doi: 10.1111/j.1467-9280.2009.02469.x

Charpentier, C. J., Iigaya, K., & O'Doherty, J. P. (2020). A Neuro-computational Account of Arbitration between Choice Imitation and Goal Emulation during Human Observational Learning. *Neuron*, *106*(4), 687–699.e7. doi: 10.1016/j.neuron.2020.02.028

Dayan, P., & Daw, N. D. (2008). Decision theory, reinforcement learning, and the brain. *Cognitive, Affective, & Behavioral Neuroscience*, *8*(4), 429–453. doi: 10.3758/CABN.8.4.429

Ewens, W. J. (2004). *Mathematical Population Genetics* (Vol. 27; S. S. Antman, J. E. Marsden, L. Sirovich, & S. Wiggins, Eds.). New York, NY: Springer. doi: 10.1007/978-0-387-21822-9

Giuliano, P., & Nunn, N. (2021). Understanding Cultural Persistence and Change. *The Review of Economic Studies*, *88*(4), 1541–1581. doi: 10.1093/restud/rdaa074

Hassabis, D., Kumaran, D., Summerfield, C., & Botvinick, M. (2017). Neuroscience-Inspired Artificial Intelligence. *Neuron*, *95*(2), 245–258. doi: 10.1016/j.neuron.2017.06.011

Henrich, J. (2017). *The Secret of Our Success*. Princeton: Princeton University Press.

Henrich, J. (2021). Cultural evolution: Is causal inference the secret of our success? *Current Biology*, *31*(8), R381–R383. doi: 10.1016/j.cub.2021.02.013

Henrich, J., & Gil-White, F. J. (2001). The evolution of prestige: freely conferred deference as a mechanism for enhancing the benefits of cultural transmission. *Evolution and Human Behavior*, *22*(3), 165–196. doi: 10.1016/S1090-5138(00)00071-4

Heyes, C. (2018). *Cognitive Gadgets: The Cultural Evolution of Thinking*. Cambridge, MA: Harvard University Press.

Heyes, C., Chater, N., & Dwyer, D. M. (2020). Sinking In: The Peripheral Baldwinisation of Human Cognition. *Trends in Cognitive Sciences*, *24*(11), 884–899. doi: 10.1016/j.tics.2020.08.006

Heyes, C., & Moore, R. (2021). *Henrich, Heyes and Tomasello on the cognitive foundations of cultural evolution* (preprint). PsyArXiv. doi: 10.31234/osf.io/a2yqr

Kleiman-Weiner, M., Sosa, F., Thompson, B., van Opheusden, B., Griffiths, T. L., Gershman, S., & Cushman, F. (2020). Downloading Culture.zip: Social learning by program induction. *Proceedings of the Cognitive Science Society*, 7.

Nowak, M. A., Sasaki, A., Taylor, C., & Fudenberg, D. (2004). Emergence of cooperation and evolutionary stability in finite populations. *Nature*, *428*(6983), 646–650. doi: 10.1038/nature02414

Rogers, A. R. (1988). Does Biology Constrain Culture? *American Anthropologist*, *90*(4), 819. doi: 10.1525/aa.1988.90.4.02a00030

Schuster, P., & Sigmund, K. (1983). Replicator dynamics. *Journal of Theoretical Biology*, *100*(3), 533–538. doi: 10.1016/0022-5193(83)90445-9

Shafto, P., Goodman, N. D., & Griffiths, T. L. (2014). A rational account of pedagogical reasoning: Teaching by, and learning from, examples. *Cognitive Psychology*, *71*, 55–89. doi: 10.1016/j.cogpsych.2013.12.004

Simpson, G. G. (1953). The Baldwin Effect. *Evolution*, *7*(2), 110–117. doi: 10.2307/2405746

Taylor, P. D., & Jonker, L. B. (1978). Evolutionary stable strategies and game dynamics. *Mathematical Biosciences*, *40*(1), 145–156. doi: 10.1016/0025-5564(78)90077-9

Tomasello, M., Kruger, A., & Ratner, H. (1993). Cultural Learning. *Behavioral and Brain Sciences*, *16*, 495–511. doi: 10.1017/S0140525X0003123X

Vélez, N., & Gweon, H. (2021). Learning from other minds: an optimistic critique of reinforcement learning models of social learning. *Current Opinion in Behavioral Sciences*, *38*, 110–115. doi: 10.1016/j.cobeha.2021.01.006

Wu, C. M., Vélez, N., & Cushman, F. A. (2022). Representational Exchange in Human Social Learning: Balancing Efficiency and Flexibility. In C. M. Wu, E. Schulz, & I. Cogliati Dezza (Eds.), *The Drive for Knowledge: The Science of Human Information Seeking* (pp. 169–192). Cambridge: Cambridge University Press. doi: 10.1017/9781009026949.009