**Title**
The Emergence of Cultural Attractors: An Agent-Based Model of Collective Perceptual Alignment

**Permalink**
https://escholarship.org/uc/item/4qc3w4pw

**Journal**
Proceedings of the Annual Meeting of the Cognitive Science Society, 43(43)

**ISSN**
1069-7977

**Authors**
Falandays, James Benjamin
Smaldino, Paul

**Publication Date**
2021

Peer reviewed

# The Emergence of Cultural Attractors: An Agent-Based Model of Collective Cognitive Alignment

**J. Ben Falandays & Paul E. Smaldino**


## Abstract

Cultural attractor landscapes describe the time-evolution of cultural variants over transmission events. When variants sit at a local minimum of a stable attractor landscape, there will be no cumulative error over transmissions, laying the foundation for cumulative culture. But because cultural attractors are emergent products of dynamic populations of *cognitive* landscapes, which are in turn emergent products of individual experience within a culture, stable cultural attractor landscapes cannot be taken for granted. Yet, little is known about how cultural attractors form or stabilize. We present an agent-based model of cultural attractor dynamics, which adapts a cognitive model of unsupervised learning of phoneme categories in individual learners to a multi-agent, sociocultural setting wherein individual learners provide the training input to each other. We find that constraints at the level of cognition, development, and demographic structure determine the tendency for populations to self-organize into and dynamically stablilize a cultural attractor landscape.


**Keywords:** cultural evolution, cultural attraction, categorization, symbolic cognition, agent-based modeling

## Introduction

In research on cultural evolution, there has been a historical and theoretical divide between approaches that emphasize information preservation, and those that emphasize information transformation (Buskell, 2017). The preservative approach can be identified with Darwinian selectionist theories of culture, which tend to focus on the fitness consequences of cultural phenotypes, and to treat transmission as analogous to biological inheritance with noise (Boyd & Richerson, 2005). The transformative approach can be identified with Cultural Attraction Theory (CAT), which emphasizes that cultural transmission is not a simple copying process, but rather an active construction process in which cognitive biases can introduce error (Sperber, 1996). The distribution of cognitive biases in a population can be thought of as comprising a cultural attractor landscape, whereby some transformations of artifacts become more likely than others.

The scientific consensus now appears to be that there is room, and indeed need, for both approaches (Buskell, 2017). Henrich, Boyd, and Richerson (2008) pointed out that Darwinian models of cultural selection are useful precisely *because* of the existence of cultural attractors: when attractors are present, cultural variants will cluster in predictable ways such that they can effectively be approximated as discrete traits. This stance puts them in agreement with Claidière, Scott-Phillips, and Sperber (2014), who argued that perfect replication is a special case of attraction: when cultural variants sit at local minima of a stable attractor landscape, there will be no cumulative transformation of the variant over repeated transmissions, allowing pure selection to dominate. Thus, both schools of thought acknowledge that cultural attraction effects are important factors for cumulative culture.

It has been suggested that cultural attractors are not actually explanations of anything, but rather something to be explained (Scott-Phillips, Blancke, & Heintz, 2018). Yet surprisingly little attention has been paid to actually explaining how cultural attractors form, stabilize, or change over time. The few computational models of cultural attraction that exist instead treat attractors as if they are independent, pre-existing, stable entities with their own causal powers (Claidière & Sperber, 2007; Rafał, 2018; Acerbi & Mesoudi, 2015). In these models, the primary mechanism invoked is that the likelihood of a cultural variant (i.e. a behavior, artifact) being reproduced in the next generation is dependent on its proximity to an attractor point. As such, these models are little more than selectionist accounts in new clothes, with the additional downside that the forces of selection have now been abstracted away to a different level of explanation. Because cultural attractors involve the influence of memories and biases that are themselves culturally learned, the landscape of cultural attractors must necessarily emerge from dynamic cultural processes.

Culture is always changing. Therefore, any apparently stable cultural attractor landscape is only *meta*-stable, constituted by a shifting population of many individuals with (at least partially) plastic cognitive landscapes. A cognitive landscape, in turn, refers to a particular system for parsing the sensory world, storing information, and generating behaviors, which can be adjusted through learning/development. As cultural artifacts are being produced, they influence the cognitive landscapes of individuals, which can alter the cultural attractor landscape at the population level, etc. Furthermore, there is overwhelming evidence that cognitive processes are massively interactive and dynamic (Falandays, Batzloff, Spevack, & Spivey, 2020), meaning cognitive attractor landscapes are in constant flux over real time and developmental time (Smith & Thelen, 2003). Viewed in light of this constant culture-cognition feedback loop, the existence of a stable cultural attractor landscape cannot be taken for granted.

While there exist several computational models of category formation in groups (Ke, Minett, Au, & Wang, 2002; Steels & Belpaeme, 2005; Baronchelli, Gong, Puglisi, & Loreto, 2010; Skyrms, 2010; Kirby, 2001; reviewed in Kallens, Dale, & Smaldino, 2018), this work has tended to abstract over the issue of cognitive alignment between individuals, instead assuming that signals can be reliably transmitted, regardless of the potential cognitive differences between transmitters and receivers. For example, a model of the self-organization of vowel systems in groups from De Boer (2000) makes the assumption that, with just one exposure to a previously unheard vowel, individuals can readily add a new category to their cognitive repertoire that closely matches the new observation. But since the shaping of cognitive landscapes is an ongoing, dynamical process, it is not clear when such simplifications are justified. Furthermore, simply requiring multiple exposures for individuals to acquire new signals, as in a model by Reali, Chater, and Christiansen (2018), may not do justice to the way that prior experience shapes future perception and learning. Even models that account for transmission errors, such as one by Nowak, Krakauer, and Dress (Nowak, Krakauer, & Dress, 1999), often assume the existence of a set of *possible* signals that is available to all agents. Therefore, existing models do not address how cultural attractor landscapes can get off the ground.

In this paper, we present an agent-based model that can account for the emergence and dynamical stabilization of cultural attractor landscapes. We adapt a model of unsupervised learning of phoneme categories in individual learners (Toscano & McMurray, 2010) to a multi-agent, sociocultural setting wherein individual learners provide the training input to each other. Agents attempt to use their limited cognitive resources to capture the distribution of sensory signals they observe from neighbors, then use their idiosyncratic perceptual representations to generate new signals (or more broadly understood, new actions, artifacts, etc.). Beginning from a state in which all agents possess a set of randomly distributed categories of uniform probability, under some conditions populations self-organize into identifiable, more-or-less stable clusters of signals, which reveal a cultural attractor landscape. We explore the role of innate cognitive capacities, levels of transmission error, production biases, learning periods, lifespans, and population sizes to understand the conditions that may be favorable or unfavorable for cumulative culture to emerge, via collective cognitive alignment.

For present purposes, we think of the signal clusters that form as akin to proto-linguistic units, such as a phoneme set, but they might also be taken to reflect any culturally-shared pattern of *categorical* distinctions–that is, those perceptual distinctions that allow the recognition and reproduction of sensory instances *as* members of a category. Deacon (1998) explains how such distinctions are the foundation for indexical knowledge (i.e. associations *between* categories), and in turn how *shared* categorical and indexical associations provide the foundation for symbolic communication. However, the level of categorical perception is relevant for virtually every aspect of culture, not just symbolic communication, because even seemingly simple, culturally-acquired behaviors can be "causally opaque" without relevant background knowledge, such that naive observers would be unable to segment, identify and replicate the relevant parts (Csibra & Gergely, 2011). The collective alignment of categorical perceptual distinctions therefore represents a critical, yet underexplored issue for cultural evolution.

## Model Description[1]

A schematic of the model dynamics are shown in Fig. 1. The model population consists of a network of $N$ agents arranged on a fully-connected network. Agents are tasked with communicating and categorizing stimuli from one another. Stimuli are represented as points on a two-dimensional[2] Euclidean $S \times S$ space (we used $S = 100$). Each agent $i$ possesses in memory a set of $K$ categories, where each category $j$ is defined as a two-dimensional Gaussian distribution defined by a mean $\mu_{ij}$ , standard deviation $\sigma_{ij}$, an amplitude $\phi_{ij}$ and a correlation $\rho_{ij}$ between dimensions (though for simplicity, we chose to keep $\rho$ fixed at 0). Collectively, an agent's set of categories is referred to as a mixture of Gaussians (MOG), and it is the agent's mental model of the signal space. The mean of each Gaussian represents the central tendency of the category (similar to prototypes in some theories of categorization), while smaller standard deviations represent more specific categories. The amplitude $\phi_{ij}$ represents the prior probability that a random stimuli is a member of that category, also known as the base rate of the category. Upon initialization, the mean of each category for each agent is randomly drawn from a uniform distribution in [[0, 100][0,100]], with an initial standard deviation of $\sigma_{init}$ (an innate inferential prior expectation). The amplitude of each category is initialized at $1/K$, representing the notion that all categories are initially equally probable. Each agent also possesses an age, which is the number of time steps for which it has been "alive." All agents are initialized with age 0.

Dynamics of the model occur in discrete time steps, each of which consists of two stages: communication and reproduction. In the communication stage, each agent is selected, in random order, to communicate a signal to a randomly selected partner. Each agent randomly selects one neighbor with whom to communicate (the target). The communicator selects one category from their memory by probabilistically sampling a category based on the prior probability distribution over categories, which is defined by the amplitude coefficients of each Gaussian. The communicator then generates a signal by probabilistically sampling a point stimulus from the selected category (so that the most maximally likely stimulus is the category's mean value). Communicating a signal involves two potential sources of deviation from the category's distribution as it is stored in memory. First, a "prototype bias" parameter $A$ in [0, 1] induces a bias towards the mean of the category. In production, the standard deviations of the selected category in each dimension are scaled by $1-A$. This means that as $A$ approaches 1, agents always produce the mean values of the selected category (because the scaled $\sigma$'s are 0), effectively implementing a "prototype" model in pro-

[1] The code to run this model is available as a Jupyter notebook on our Github page: https://github.com/bfalandays/CulturalAttractorDynamics

[2] Note that while we do not expect the qualitative patterns of our model to be dependent upon restricting the signal space to two dimensions, the effect of adding additional dimensions has yet to be systematically explored.

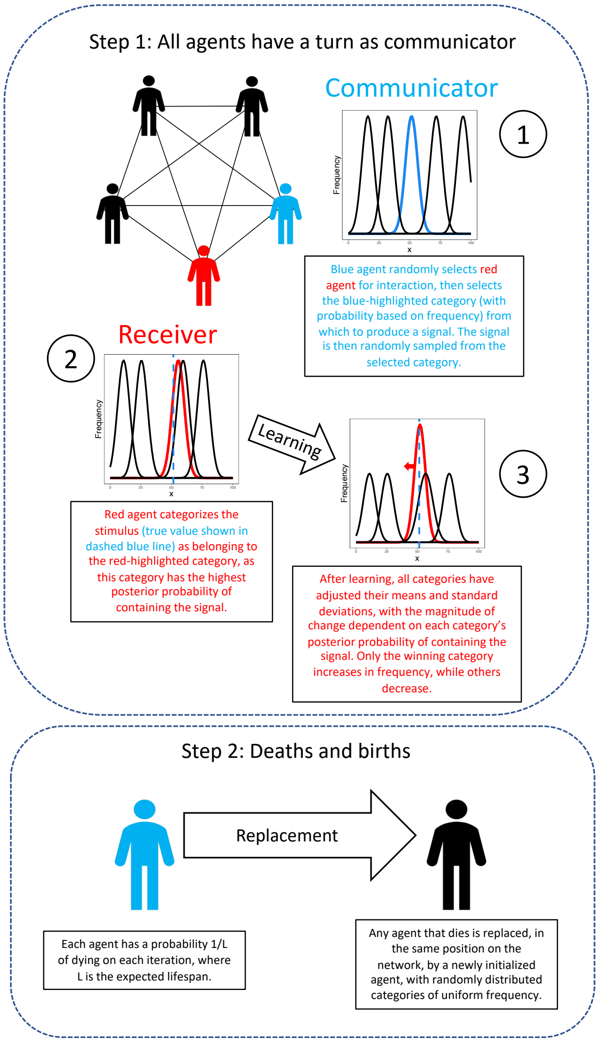


Figure 1: An illustration of the model dynamics. Note that, for ease of visualization, cognitive landscapes are illustrated here as a mixture of 1-D Gaussians, when the actual model used 2-D Gaussians.

duction. On the other hand, as $A$ approaches 0, the probability density function of possible signals approaches the full Gaussian distribution observed by the agent. The $A$ parameter is intended to capture the fact that production of a category within an individual $i$ is less variable than production of that category across observations $i$ has made of other individuals. For example, while it is easy to recognize the same word produced by many speakers with different accents, a single individual will typically produce the word with only one accent. Second, after a "raw" signal is generated by an agent, transmission noise is added, which is here implemented as Gaussian noise with $\mu = 0$ and $\sigma = W$.

Upon receiving a signal from a communicator, the target agent uses Bayesian inference to categorize the signal and adjust the parameters of their MOG representation in memory. The target first maps the signal as a member of the most likely of its own stored categories, then updates the properties of its categories to reflect this new information. Critically, updating of the frequency parameter of Gaussians occurs based on winner-take-all competition, such that only the most likely Gaussian for a signal increases in frequency, while all other decrease slightly. Because the mathematical description of the learning process is somewhat complicated, for considerations of space we refer interested readers to the learning rules as described in Toscano & McMurray (2010).

After each agent has had the opportunity to communicate (not all agents will receive a signal, and some will receive multiple signals on a given time step), the reproduction stage occurs. On each step, each agent has a probability of $1/L$ of dying, giving an expected lifespan of $L$ iterations. Any agent who dies is removed from the simulation and replaced by a new agent, initialized in the same way as agents at the beginning of each simulation. We also explored the influence of a "critical period" for learning, implemented by turning off learning for agents over $C$ iterations in age.

## Simulation Experiments

### Baseline Model: Qualitative Analysis and Visualization

In order to understand the influence of informational bottlenecks at the cognitive, developmental, and demographic levels on the tendency to self-organize into a stable cultural attractor landscape, it will be helpful to first qualitatively analyze the behavior of a baseline model. The parameters used in the baseline model are presented in Table 1, in bold font.

Fig. 2 provides a representative illustration of the behavior of the model. The model begins with all agents possessing a set of equal-frequency Gaussians, uniformly distributed throughout the signal space. Over the first 5000 time steps, we can see the beginnings of cultural attractors emerging, as nearby Gaussians begin to converge over the course of communication and learning. By 10,000 time steps, a clearly distinguishable set of tight clusters have emerged, though there remain some looser clouds of low-frequency categories, driven by new learners entering the population. At this point, the qualitative structure of the clustering pattern remains stable, but clusters continue to drift around the signal space stochastically. Some categories move too near to each other and merge[3], while new clusters emerge to fill empty spaces. In other words, the dynamics of the baseline model result in a number of discriminable, roughly-shared perceptual categories (i.e, cultural attractors) emerging in the population. However, these patterns are unstable, which would likely impact their utility as scaffolds for other cultural constructions.

[3]Note that categories at the level of agents do not merge. Instead, if two categories become too close to each other, they will compete within an agent's MOG, which can result in one category increasing in frequency while the other diminishes. On the other hand, the categories detected at the population scale, using the k-means algorithm, do not directly compete, and thus may be described as "merging" when the algorithm detects two nearby clusters at one time point, but detects only a single cluster at a subsequent time point that en-

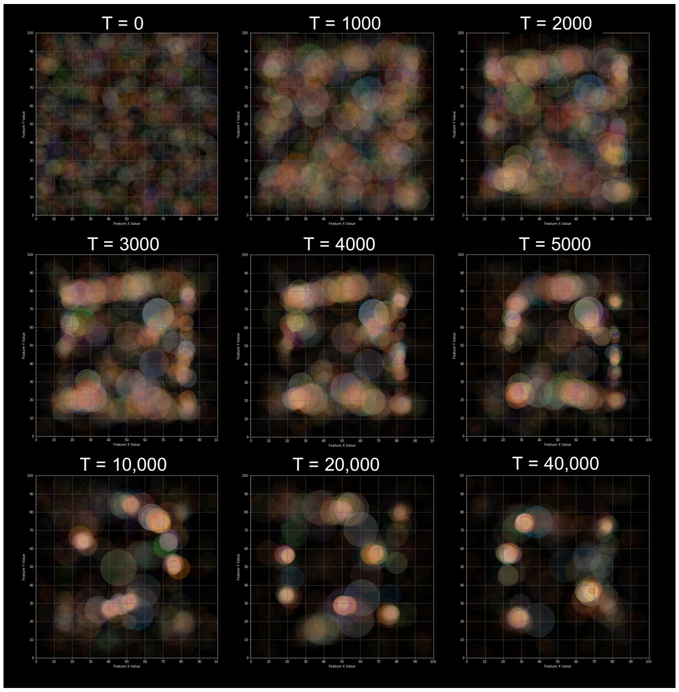


Figure 2: The states of all Gaussian across all agents in a population at nine time points of one run. Colors correspond to different agents. Each agent has multiple categories in their MOG, which correspond to individual points of a given color. The size of points is proportional to the SD of the category, and the transparency (alpha value) of the points is proportional to the frequency of the category, such that infrequent categories become more transparent.

## Outcome Measures

Our model of the emergence of cultural attractors is not a model of cultural evolution per se, due to the fact that our attractors are currently completely arbitrary and fitness neutral. Rather, it is a model of the emergence of some of the preconditions for cultural evolution to take place: shared perceptual biases (e.g., analogical categories, iconic distinctions) that allow for cultural variants to be faithfully transmitted (Brand, Mesoudi, & Smaldino, 2021). In other words, cultural attractor dynamics are necessary but not sufficient to explain cumulative culture, which requires at minimum the additional step of mapping the attractor landscape onto a fitness landscape. As a first step towards this, our goal at present was to characterize the behavior of the model in ways that plausibly relate to the potential for cumulative culture under various parameter settings. To this end, we considered a number of measures related to the 1) complexity, 2) discriminability, 3) stability, and 4) conformity (i.e. degree of inter-agent agreement) with respect to the inferred attractor landscape. We expect these outcomes may be relevant in different ways for different cultural domains or problem spaces, and hence at this level of generality we cannot yet say which outcome patterns are ”optimal” for cumulative culture.

To obtain our measures, the model was observed every 1000 time steps by generating 500 signals from each agent (using the same method as for communication). Additionally, the state of all agents' MOGs was recorded at the end of each run, in order to characterize cognitive patterns at the agent level. To characterize the emergent cultural attractor landscape at the population level, at each time slice of the data we applied the k-means algorithm. To determine the optimal value for $k$, the partition was calculated at each evaluated time point using values of $k$ ranging from 1-50. We then used the gap statistic (Tibshirani, Walther, & Hastie, 2001) to select the optimal value of $k$ at each timepoint.

Based on the obtained optimal k-means partition at each time step, we operationalized the **complexity** of the attractor landscape as the Shannon Entropy of the frequency distribution over clusters. The **discriminability** of clusters was operationalized as the silhouette coefficient for the optimal k-means partition of each sample of signals, a common metric for validating clustering schemes. This value ranges between $[-1, 1]$, with values near 1 indicating well-separated clusters.

Next, to examine the relative **stability** of the population signal distribution, we adopted a dissimilarity metric for probability distributions known as the earth mover's distance (EMD). The EMD can be understood by imagining different probability distributions as different ways of piling up an amount of dirt (or “earth”). The dissimilarity between two distributions can be thought of as the minimal cost of moving one pile of dirt—a reference distribution—such that it is transformed into a differently-shaped pile of dirt—a target distribution. Because our signal space is continuous, to compute this measure we first constructed a discrete probability distribution based on the full set of signal samples at each time point. The signal space was divided into a grid of $20 \times 20$ evenly-spaced points (each square being $5 \times 5$) and the number of observations in each square was counted, creating a 2-D histogram. We then computed the EMD between the population distribution at each timepoint $t$ to the same population at time $t$ - 1 (therefore there is no measure taken at time 0). This provides a measure of the change in the population distribution over the time between each evaluated timepoint (the model was evaluated every 1000 timesteps).

Finally, to examine the **conformity** in cognitive biases across agents, we computed the average dissimilarity of the distribution of signals generated by an individual agent to the distribution generated from the rest of the population. We will refer to this as “nonconformity,” because this term captures the notion of divergence from a population norm. To compute this metric, we again used the EMD described above. At each evaluated timepoint, a 2-D histogram was constructed from the signal samples from each individual agent $i$ in a population of size $N$, and was compared to another histogram was constructed from the signal samples corresponding to every agent *besides* the focal agent (similar to the “jackknife”

compasses the former two.

resampling technique). Finally, we took the average of these values across agents, which provides a measure of the relative conformity vs. idiosyncracy, or generalization vs. specialization, in a population.

**Data Normalization** In order to understand the impact of each informational bottleneck at the cognitive, developmental, or demographic scale, we varied parameters one at a time from the parameter settings in the baseline model. The full set of parameters explored are presented in Table 1. Learning rates were held fixed across runs. In order to summarize the results, we first standardized all independent and dependent variables to allow for comparison. We next extracted the mean value for each outcome measure, over the second half of each of run (100 runs were conducted per parameter value). Then, all parameters were entered as predictors into a linear regression, treating each outcome measure as the dependent variable in turn. Finally, we extracted the regression coefficients for each parameter with respect to each outcome measure. These results are plotted in Fig. 3, which allows for a rough comparison of the impact of each parameter on each outcome measure. While this is a necessary simplification for purposes of space, it should be noted that several of our parameters show non-linear, threshold-like effects with respect to some outcome measures, which are of course not captured by linear regression, and will require more fine-grained analysis the future.

Table 1: Variable model parameters. The values used in the baseline model are presented in bold font.

| Parameter | Values Explored |
|---|---|
| $K$ (MOG size) | 10, **20**, 30 |
| $\sigma_{init}$ (Init. cat. S.D.) | 1, **5**, 10 |
| $A$ (Prototype bias) | 0, .25, .75, **1** |
| $W$ (Trans. noise) | **0**, 3, 10 |
| $C$ (Crit. period len.) | 2500, 5000, 10000, 20000, **40000** |
| $L$ (Exp. lifespan) | 5000, **10000**, 15000 |
| $N$ (Pop. size) | 10, 25, **50**, 100, 200 |

## Results and Discussion

To unpack our initial findings, we proceed upwards from the scale of cognitive constraints, to the level of development, and finally to the level of demographic constraints. Beginning at the cognitive level, the first relevant parameter is $K$, the number of Gaussians available in each agents MOG, which corresponds to memory resources. $K$ shows an intuitive positive relationship with complexity–when agents have greater memory resources, populations sustain more complex repertoires. We also observe greater stability and conformity as $K$ increases. This is due to the fact that, as agents have more memory resources available, they are more quickly and completely able to attune to the population distribution of signals. While fitness consequences are not currently represented in our model, we expect that increased memory resources come at a greater metabolic cost. As such, in future explorations this parameter can allow for examining the co-evolution of culture and the brain. This finding can also complements previous work showing how neural and cognitive limitations influence collective categorical alignment, such as the way that the just-noticeable-difference in human color perception can be sufficient to trigger cross-cultural patterns of color categorization (Baronchelli et al., 2010; Puglisi, Baronchelli, & Loreto, 2008; Gong, Gao, Wang, & Shuai, 2019).

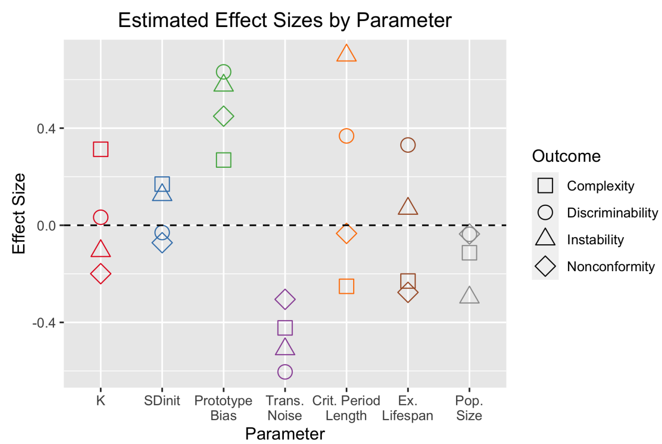


Figure 3: The estimated effect size of each IV on each DV based on linear regression, after standardizing all variables.

Another constraint at this level relates to the fact that a only a finite number of signals will be observed by any agent, introducing a role for cognitive "priors" to guide inference: because learners are not exposed to all possible signals, repeated inference processes can lead population distributions to converge towards prior expectations (Kirby, Cornish, & Smith, 2008). In our model, priors are implemented with the $\sigma_{init}$ parameter (as well as the initial locations of Gaussians, though this is currently random rather than inherited), which determines the expected variability of categories. We find that increasing $\sigma_{init}$ from the baseline value of 5 to a value of 10 has little impact on final outcomes, while decreasing $\sigma_{init}$ from 5 to 1 results in a collapse of the cultural repertoire towards a single, central attractor (though this non-linearity is not well revealed by Fig. 3). This threshold effect occurs because when categories are initially very tight, they act as extremely powerful *cognitive* attractors, heavily concentrating activation (on the basis of input) to just one category at a time. These extreme competition effects eventually result in only one major attractor in the global space.

The next set of parameters can be grouped as sources of transmission noise. We considered both general transmission noise ($W$) and noise introduced by production biases ($A$). Note that the prototype bias parameter, $A$, is functionally the inverse of the noise parameter $W$(greater values of $A$ *reduce* variability around the mean, while greater values of $W$ *in-*

*crease* variability), which is consistent with the inverse patterns revealed in Fig. 3. In our model, we find that sources of noise modulate a trade-off between complexity and discriminability, or stability and conformity. By limiting the fidelity of transmission, low-moderate levels of noise make signalling less precise, which slows learning rates and thereby stabilizes the global pattern. This helps to promote cognitive conformity across individuals. However, too much noise limits the discriminability of categories, and therefore the complexity of the population-level clustering scheme–when categories are highly variable, fewer of them can be maintained in the same space. This result complements a point made in the literature on "stochastic resonance" that some noise is crucial for the stability and sensitivity of complex dynamical systems (Wiesenfeld & Moss, 1995; Turner & Smaldino, 2018).

Proceeding to the developmental scale, we next consider the effect of a critical period for learning ($C$) as well as the expected lifespan of individuals ($L$). In the case of culture, where meaningful distinctions may be fine grained and rapidly changing, investment in learning is critical. However, time spent learning is costly and detracts from reproduction, therefore longer learning times should be selected against, all else being equal (Hurford, 1991). This means that a cultural repertoire is constrained to being learnable, but the existence of the repertoire may itself drive selection for greater investment in learning, which then allows for more complex culture, and so on. Furthermore, individuals must live long enough to transmit what they have learned, else the culture will simplify or dissipate entirely.

We find that the presence of a critical period substantially enhances the stability of the attractor landscape over time. Shorter learning times also can increase the complexity of the landscape (seen as a negative effect of critical period *length* on complexity in Fig. 3). However, if learning times are *too* short (e.g. 2500 time steps in our model), this leads to a potential cost in terms of decreased discriminability of categories and increased non-conformity, as individuals are unable to converge to the dominant global pattern before the critical period closes (note that this non-linearity is not well revealed by Fig. 3). On the other hand, when learning proceeds over the entire lifespan, we find that longer lifespans result in a decrease in complexity at the population-level, which is attributable to competition between categories–the more time spent learning, the more categories that will be irrecoverably suppressed. With fewer categories, those that remain then become more discriminable. These findings add to a growing body of work on the evolution of critical or sensitive periods (Frankenhuis & Walasek, 2020) by suggesting that, in the case of culture, long learning times may not be selected against only due to metabolic costs, but also by virtue of their role in stabilizing a cultural attractor landscape.

Finally, at the demographic level, we considered the influence of population size $N$. Population size and/or density are considered important factors for cultural evolution because these parameters constrain the number of neighbors from which an individual may learn, which can limit the complexity sustainable by a population (Henrich, 2004). In our model, we find that smaller populations are able to maintain slightly more complex distributions, but that distributions are substantially more stable within large populations. This is the inverse pattern typically pointed to in the literature (larger populations are thought to sustain more complex repertoires), but we again point out that our cultural attractor landscapes represent shared *perceptual, categorical* distinctions, whereas the cultural variants referred to in the literature on population size/density are generally combinatorial productions scaffolded by iconic distinctions (i.e. tools). As such, this finding is not contradictory to previous work, and instead suggests interesting relationships between the complexity of perceptual category repertoires, and the complexity of the combinatorial repertoires they support.

## Conclusion and Future Directions

While much of the work on the cultural evolution of arbitrary conventions has focused on the mapping between signals and meanings, taking for granted a shared set of perceptual categories as a starting point, we have tried to show that this is not trivial issue. It is only when factors at multiple levels of analysis dynamically constrain the degrees of freedom in cultural transmission in complementary ways that a shared set of cultural attractors may form and stabilize, which can then provide the foundation for selection on symbolic systems (Jablonka & Lamb, 2014; Deacon, 1998). Viewing cultural attractor landscapes as a complex system of interacting constraints allows for straightforward integration of CAT with Darwinian selectionist accounts: fitness-based selection effects can be understood as yet another constraint promoting the formation of statistical attractor points. As such, we believe this model may be useful for researchers interested in the co-evolution of innate cognitive biases, developmental tendencies, and demographic structure with culture.

Our model can be straightforwardly extended to incorporate biological inheritance of cognitive priors and/or developmental hyper-parameters, as well as to include fitness constraints, by placing our agents into any type of evolutionary or communicative game. As such, our model can be used to simulate the operation of selectionist dynamics and transformative dynamics simultaneously. A next step will be to allow agents to generate *sequences* of signals, allowing us to examine the entanglement between perceptual and combinatorial/syntactic cognition in cultural attractor dynamics.

## References

Acerbi, A., & Mesoudi, A. (2015). If we are all cultural darwinians what's the fuss about? clarifying recent disagreements in the field of cultural evolution. *Biology & philosophy*, *30*(4), 481–503.

Baronchelli, A., Gong, T., Puglisi, A., & Loreto, V. (2010). Modeling the emergence of universality in color naming patterns. *Proceedings of the National Academy of Sciences*, *107*(6), 2403–2407.

Boyd, R., & Richerson, P. J. (2005). *The origin and evolution of cultures*. Oxford University Press.

Brand, C., Mesoudi, A., & Smaldino, P. (2021). Analogy as a catalyst for cumulative cultural evolution. *Trends in Cognitive Sciences*.

Buskell, A. (2017). What are cultural attractors? *Biology & Philosophy*, *32*(3), 377–394.

Claidière, N., Scott-Phillips, T. C., & Sperber, D. (2014). How darwinian is cultural evolution? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1642), 20130368.

Claidière, N., & Sperber, D. (2007). The role of attraction in cultural evolution. *Journal of Cognition and Culture*, *7*(1-2), 89–111.

Csibra, G., & Gergely, G. (2011). Natural pedagogy as evolutionary adaptation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1567), 1149–1157.

Deacon, T. W. (1998). *The symbolic species: The co-evolution of language and the brain* (No. 202). WW Norton & Company.

De Boer, B. (2000). Self-organization in vowel systems. *Journal of phonetics*, *28*(4), 441–465.

Falandays, J. B., Batzloff, B. J., Spevack, S. C., & Spivey, M. J. (2020). Interactionism in language: from neural networks to bodies to dyads. *Language, Cognition and Neuroscience*, *35*(5), 543–558.

Frankenhuis, W. E., & Walasek, N. (2020). Modeling the evolution of sensitive periods. *Developmental cognitive neuroscience*, *41*, 100715.

Gong, T., Gao, H., Wang, Z., & Shuai, L. (2019). Perceptual constraints on colours induce the universality of linguistic colour categorisation. *Scientific reports*, *9*(1), 1–10.

Henrich, J. (2004). Demography and cultural evolution: how adaptive cultural processes can produce maladaptive losses: the tasmanian case. *American antiquity*, 197–214.

Henrich, J., Boyd, R., & Richerson, P. J. (2008). Five misunderstandings about cultural evolution. *Human Nature*, *19*(2), 119–137.

Hurford, J. R. (1991). The evolution of the critical period for language acquisition. *Cognition*, *40*(3), 159–201.

Jablonka, E., & Lamb, M. J. (2014). *Evolution in four dimensions, revised edition: Genetic, epigenetic, behavioral, and symbolic variation in the history of life*. MIT press.

Kallens, P. A. C., Dale, R., & Smaldino, P. E. (2018). Cultural evolution of categorization. *Cognitive Systems Research*, *52*, 765–774.

Ke, J., Minett, J. W., Au, C.-P., & Wang, W. S.-Y. (2002). Self-organization and selection in the emergence of vocabulary. *Complexity*, *7*(3), 41–54.

Kirby, S. (2001). Spontaneous evolution of linguistic structure-an iterated learning model of the emergence of regularity and irregularity. *IEEE Transactions on Evolutionary Computation*, *5*(2), 102–110.

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, *105*(31), 10681–10686.

Nowak, M. A., Krakauer, D. C., & Dress, A. (1999). An error limit for the evolution of language. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *266*(1433), 2131–2136.

Puglisi, A., Baronchelli, A., & Loreto, V. (2008). Cultural route to the emergence of linguistic categories. *Proceedings of the National Academy of Sciences*, *105*(23), 7936–7940.

Rafał, M. (2018). The relationship between the accuracy of cultural transmission and the strength of cultural attractors. *Social Evolution & History*, *17*(2).

Reali, F., Chater, N., & Christiansen, M. H. (2018). Simpler grammar, larger vocabulary: How population size affects language. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1871), 20172586.

Scott-Phillips, T., Blancke, S., & Heintz, C. (2018). Four misunderstandings about cultural attraction. *Evolutionary Anthropology: Issues, News, and Reviews*, *27*(4), 162–173.

Skyrms, B. (2010). The flow of information in signaling games. *Philosophical Studies*, *147*(1), 155–165.

Smith, L. B., & Thelen, E. (2003). Development as a dynamic system. *Trends in cognitive sciences*, *7*(8), 343–348.

Sperber, D. (1996). Explaining culture: A naturalistic approach. *Cambridge, MA: Cambridge*.

Steels, L., & Belpaeme, T. (2005). Coordinating perceptually grounded categories through language: A case study for colour.

Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *63*(2), 411–423.

Toscano, J. C., & McMurray, B. (2010). Cue integration with categories: Weighting acoustic cues in speech using unsupervised learning and distributional statistics. *Cognitive science*, *34*(3), 434–464.

Turner, M. A., & Smaldino, P. E. (2018). Paths to polarization: How extreme views, miscommunication, and random chance drive opinion dynamics. *Complexity*, *2018*, 2740959.

Wiesenfeld, K., & Moss, F. (1995). Stochastic resonance and the benefits of noise: from ice ages to crayfish and squids. *Nature*, *373*(6509), 33–36.