COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 46 (2022) e13183

ISSN: 1551-6709 online
DOI: 10.1111/cogs.13183

The Emergence of Cultural Attractors: How Dynamic Populations of Learners Achieve Collective Cognitive Alignment

J. Benjamin Falandays,[a,b] Paul E. Smaldino[a]

[a]*Department of Cognitive and Information Sciences, University of California, Merced, United States*
[b]*Department of Cognitive Linguistic and Psychological Sciences, Brown University*

Received 3 August 2021; received in revised form 24 June 2022; accepted 18 July 2022

Abstract

When a population exhibits collective cognitive alignment, such that group members tend to perceive, remember, and reproduce information in similar ways, the features of socially transmitted variants (i.e., artifacts, behaviors) may converge over time towards culture-specific equilibria points, often called cultural attractors. Because cognition may be plastic, shaped through experience with the cultural products of others, collective cognitive alignment and stable cultural attractors cannot always be taken for granted, but little is known about how these patterns first emerge and stabilize in initially uncoordinated populations. We propose that stable cultural attractors can emerge from general principles of human categorization and communication. We present a model of cultural attractor dynamics, which extends a model of unsupervised category learning in individuals to a multiagent setting wherein learners provide the training input to each other. Agents in our populations spontaneously align their cognitive category structures, producing emergent cultural attractor points. We highlight three interesting behaviors exhibited by our model: (1) noise enhances the stability of cultural category structures; (2) short 'critical' periods of learning early in life enhance stability; and (3) larger populations produce more stable but less complex attractor landscapes, and cliquish network structure can mitigate the latter effect. These results may shed light on how collective cognitive alignment is achieved in the absence of shared, innate cognitive attractors, which we suggest is important to the capacity for cumulative cultural evolution.

Keywords: Agent-based modeling; Cultural evolution; Cultural attraction; Categorization; Symbolic cognition

Correspondence should be sent to J. Benjamin Falandays, Department of Cognitive and Information Sciences, University of California, Merced, Merced, CA 95343, USA. E-mail: jfalandays@ucmerced.edu

1. Introduction

All human groups possess group-specific behavioral repertoires involving cultural variants—things such as tools, linguistic behavior, social norms, religious beliefs, and artistic styles. As cultural variants are observed and copied, they are liable to change over time as a result of noise, errors, and biases in both transmission and interpretation. However, even in the absence of strong selection for specific outcomes, cultural variants may nevertheless converge over successive transmission events towards culture-specific 'attractor' points (Sperber, 1996). This effect can be attributed, at least in part, to the fact that individuals within a cultural group often share similar cognitive biases, such that they tend to perceive, remember, and reproduce information in consistent ways (Heyes, 2018). Without this "cognitive alignment," cultural transmission would be far less reliable, and the potential for cumulative cultural evolution would be limited.

But how does cognitive alignment first emerge in initially uncoordinated, dynamic populations? Current models of cultural evolution usually take cognitive alignment as given. This is implicitly the case in most models based on the mathematics of population genetics or epidemiology, which assume high-fidelity transmission (Acerbi, Mesoudi, & Smolla, 2020; Boyd & Richerson, 1988; Cavalli-Sforza & Feldman, 1981; Mesoudi, 2021), and explicitly the case in most models of cultural attraction, in which any factors of attraction are assumed to be both stable and universally shared throughout the population (Acerbi, Charbonneau, Miton, & Scott-Phillips, 2019; Claidière and Sperber, 2007; Mesoudi, 2021; Rafał, 2018). There are, to our knowledge, no models that demonstrate how attractors arise. This is an important gap in theory in light of the many cases where culture depends on cognitive biases that are themselves socially acquired (Heyes, 2018; Karmiloff-Smith, 1994), and therefore not guaranteed. Within shifting populations of cognitively plastic individuals, cognitive alignment may need to be actively and continuously maintained in order for cultural knowledge to be successfully preserved across generations.

In this paper, we develop an agent-based model of the emergence and maintenance of collective cognitive alignment in dynamic populations. Our model adapts an existing model of unsupervised learning of phoneme categories in individual learners (Toscano and McMurray, 2010) to a multiagent, sociocultural setting wherein individual language 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. 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 signal clusters, which constitute an identifiable set of cultural attractors. These cultural attractors may be thought of as akin to protolinguistic units, such as a set of phonemes, but also may be taken to represent any culturally shared repertoire of categories or behaviors. We explore the role of various innate cognitive constraints, levels of transmission error, learning periods, lifespans, population sizes, and network structures to understand when the population-level structure may emerge, what properties it is likely to have, and how stable it is.

Our explorations with this model suggest that achieving and maintaining cognitive alignment may depend upon a finely tuned balance of factors at the levels of cognition, development, and demographic structure. We highlight three interesting and potentially counterintuitive behaviors exhibited by our model that are not accounted for in other models of cultural evolution: First, we find that some noise is beneficial to stabilizing cognitive alignment. Second, we find that long learning times may destabilize and limit the complexity of cultural repertoires, while critical or sensitive periods of learning enhance stability. Third, we find that larger populations develop less complex, but more stable patterns of alignment and that this effect can be moderated by the network structure. These results suggest that additional complexity may be needed in models of cultural evolution to adequately understand how human-level culture can get off the ground and develop. We conclude by highlighting several ways that our model may be extended to complement existing models of cultural evolution and gene-culture co-evolution.

1.1. Why we need more models of cultural attraction

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 treat transmission as analogous to biological inheritance with noise (Boyd & Richerson, 1988; Cavalli-Sforza & Feldman, 1973, 1981; Dawkins, 1976; Smaldino, 2014). This often reflects a modeling simplification rather than a deep assumption about the intrinsic nature of cultural transmission, as simplifying assumptions are needed to advance theory (Healy, 2017; Smaldino, 2017). However, some researchers have argued that high-fidelity copying is more than just a simplifying assumption but in fact one of the keys to cumulative cultural evolution (H. M. Lewis & Laland, 2012), bolstering this claim with evidence from cross-species studies showing that humans are exceptional- or even *over*-imitators, often copying observed actions even when they are causally irrelevant to an outcome (Horner & Whiten, 2005; Hoehl et al., 2019). In sum, the idea of high-fidelity copying has played a substantial role in explanations of the human capacity for cumulative cultural evolution.

The transformative approach, in contrast, can be identified with cultural attractor theory (CAT), which emphasizes the fact that individuals have potentially idiosyncratic cognitive biases in how they process and reconstruct cultural variants, such that cultural transmission may not conform to the predictions of a gene-like inheritance system (Claidière, Scott-Phillips, & Sperber, 2014; Sperber, 1996; Scott-Phillips, Blancke, & Heintz, 2018). The distribution of cognitive biases in a population can be thought of as comprising a "cultural attractor landscape," whereby some transformations of variants are more likely than others. An early example of this phenomenon is Bartlett's classic 'War of the Ghosts' study (Bartlett & Bartlett, 1932) , in which English participants read a Native American (Chinook) folktale and then, after various time delays, attempted to recall the content. Bartlett found that those story

elements that were inconsistent with the 'cultural schema' of the participants (i.e., the narrative patterns with which they were familiar) tended to be forgotten or transformed into more familiar forms, especially as the time delay increased. When culture-specific cognitive biases of this kind are applied iteratively in social transmission, variants may converge towards group-specific equilibria points in the space of possible features, known as cultural attractors. This phenomenon has been demonstrated with transmission chain studies using images (Bartlett & Bartlett, 1932), event descriptions (Mesoudi & Whiten, 2004), music (Ravignani, Delgado, & Kirby, 2016), grammars (Kirby, Cornish, & Smith, 2008), tools (Thompson & Griffiths, 2021), and function concepts (Kalish, Griffiths, & Lewandowsky, 2007); for a review see Miton and Charbonneau (2018).

It is increasingly recognized that there is room, and indeed need, for both approaches (Buskell, 2017; Mesoudi, 2021). Yet, in spite of this nominal consilience, little traction has been gained towards developing a theory that integrates both preservative and transformative factors in cultural evolution. For example, a 2015 review by Acerbi and Mesoudi reported only one known empirical study designed to address both selection and attraction effects simultaneously. This represents a crucial missing link in the literature, given that these two approaches do not deal with neatly separable scales of analysis (Wimsatt, 1972).

A major barrier towards the fruitful interaction between these two perspectives is a dearth of formal models of cultural attraction. Cultural attractors are said to be statistical abstractions, and therefore to be the primary phenomenon in need of explanation (Scott-Phillips et al., 2018), yet there are no mechanistic models of how cultural attractors form, stabilize, or change over time. The few computational models of cultural attraction that exist instead make the assumption of pre-existing, stable cultural attractor points (Acerbi et al., 2019; Acerbi, Charbonneau, Miton, & Scott-Phillips, 2021; Claidière and Sperber, 2007; Mesoudi, 2021; Rafał, 2018). The cognitive or ecological forces that determine attractor points are assumed to be shared across members of a population from the outset and stable across generations of individuals. While these models have been useful in showing how the presence of cultural attractors can influence the distribution of cultural variants over time, they are agnostic with respect to how cultural attractors initially form or potentially change over time.

In this paper, we model cultural attractors as arising from the collective alignment of cognitive landscapes within a population (see Fig. 1 and Table 1). A cognitive landscape refers to a particular way of parsing the sensory world, storing information, and generating behaviors, which determines the transformation one individual will apply when reproducing a cultural variant from a model. When many individuals within a population develop aligned cognitive landscapes, social transmission becomes more reliable because many different individuals apply convergent transformations to information upon reproduction, allowing cultural variants to cluster into distinct types. In some cases, the alignment of cognitive landscapes in a population may be the result of highly canalized developmental trajectories driven by genetic evolution. However, many important aspects of human culture rely on cognitive biases that are themselves socially transmitted. As such, our goal in this paper is to offer an account of

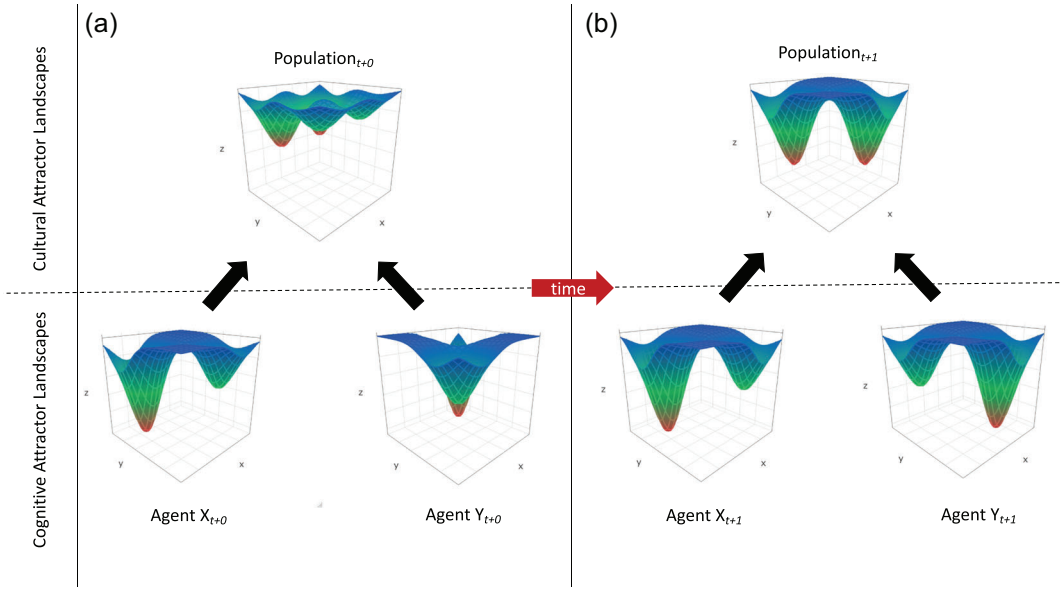

Fig. 1. A simplified illustration of the feedback loop between cognitive and cultural attractor landscapes. An attractor landscape, generally speaking, is a function describing the rate and direction of change of some variable(s), which can be visualized as a hypersurface. Valleys in an attractor landscape correspond to attractors—local equilibria towards which outputs converge over time—with the strength of attraction represented by the steepness of the valley. Here the *x*- and *y*-axes may represent any two dimensions of variability in a cultural variant (e.g., length and width of an arrow head; speech cues such as voice-onset-time and fundamental frequency). A cognitive attractor landscape (lower panels) gives the expected transformation that one individual will apply when attempting to reproduce a cultural variant from an observation. In the lower panels, we show the cognitive landscapes of two individuals, each containing two cognitive attractors of differing location and strength. Multiple cognitive landscapes can be averaged to produce a cultural attractor landscape (upper panels) that gives the expected change in a distribution of cultural variants over the course of multiple transmissions within a population. Panel A shows a situation in which two individuals have disaligned cognitive landscapes, resulting in a rugged cultural landscape with four weak cultural attractors. Panel B shows the same two individuals at a later time, with agent Y having more closely aligned their cognitive landscape to individual X, resulting in a smoother cultural landscape with two strong cultural attractors.

the emergence of cultural attractors specifically in cases for which there are not yet shared, innate cognitive attractors.

In order to advance this argument, we first present evidence that cultural attractor theory supports a Darwinian view of cultural evolution. Next, we consider several possible mechanisms of attraction, including evolved cognitive biases and shared ecological constraints, but emphasize the importance of collective cognitive alignment through enculturation. Then, we describe how culturally shared cognitive biases could emerge in a cognitively dynamic population. Finally, we support our theory with an agent-based model that can account for the emergence of cultural attractors through the lower level interactions among cognitive agents, without appealing to selectionist principles.

Table 1
Key concepts

Key Concept	Definition
Cultural variants	Behaviors or artifacts generated by individuals in a cultural population.
Cognitive landscape	A cognitive function giving the probability of different outputs (e.g., neural states, behaviors, cultural variants produced) for an individual, given some range of inputs. Represents the cumulative effects of sensation, perception, memory, attention, motor control, and any other cognitive processes that shape how an individual responds to stimuli and generates new behaviors.
Cognitive attractors	Local minima in a cognitive landscape, corresponding to outputs that are more likely for an individual in general, or more likely in response to some particular input.
Cultural landscape	A function describing the probability of observing different cultural variants at a population level. Represents the aggregate result of a population of cognitive landscapes, plus patterns of social interaction and any ecological factors that influence the observation and reproduction of cultural variants.
Cultural attractors	Local minima in a cultural landscape, corresponding to high-probability variants for a population in general, or semistable equilibria towards which transformations converge given an initial distribution of variants.
Culture-cognition feedback loop	The co-evolution of a cultural landscape with a population of cognitive landscapes. As each generation of individuals learns from exposure to a distribution of cultural variants, this may result in a change to the set of cognitive landscapes, in turn producing a new distribution of cultural variants in the next generation, and so on.
Collective cognitive alignment	The convergence of cognitive landscapes within a cultural group in the absence of innately shared cognitive attractors, such that group members tend to perceive, remember, and reconstruct information in convergent ways.

1.1.1. The role of cultural attractors in Darwinian cultural evolution and information transfer

Cultural attraction theory is often framed as a critique or qualification of Darwinian selectionist models of cultural evolution, in which cultural variants are often modeled as discrete units that are more or less faithfully transmitted (similar to "memes" as described by Dawkins, 1976). But even as CAT challenges the assumption of high-fidelity copying, it simultaneously describes the conditions under which this assumption may be justified: when variants have converged to a cultural attractor point, such that subsequent transmission events no longer incur systematic deviations from a model. Prominent researchers associated with both Darwinian and CAT research camps have pointed to this complementarity between their approaches. Henrich, Boyd, and Richerson (2008) explain that Darwinian models of cultural selection are useful precisely *because* of the existence of cultural attractors (see also Henrich and Boyd (2002)): in their model, so long as there is more than one attractor present in space of cultural variation, transmission errors will be corrected to some extent and cultural phenotypes will cluster such that they can effectively be approximated as discrete traits. This stance puts Henrich et al. (2008) in agreement with Claidiére, 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 cultural attractor landscape, there will be no bias in the transformation of the variant over repeated transmissions, allowing pure selection to dominate (see also Claidière and Sperber (2007)). In this way, the existence of cultural attractors lays the foundation for cumulative culture.

Another way to understand the important role of cultural attractors in Darwinian cultural evolution relates to the capacity for information transfer. Consider that all information transfer presupposes a particular reference frame for distinguishing signal from noise in a continuous physical channel (Fields & Levin, 2020; Von Uexküll, 1934). All information transfer is "transformative" to an extent, in that any sender must apply some function for encoding messages into physical signals, and any receiver must apply some function for decoding messages *from* signals, with both processes inevitably subject to noise, however small. However, information can be preserved when senders and receivers share a reference frame, such that the transformations applied in encoding/decoding are convergent. For example, binary digital signals may be represented as voltages near 0 for *off*, and near 5 for *on*, perhaps using a simple threshold function (i.e., values below 2.5 V are treated as *off*, and values above 2.5 V are treated as *on*). Given noise, a sender may produce a voltage of $+1$ V or -1 V on different instances when trying to communicate an *off* message, but in both cases the signal will be compressed into an *off* message by a receiver (with the same reference frame) before passing the message along again, which prevents the accumulation of noise. However, if senders and receivers do not define the same set of signals over the communication channel and/or encode messages into signals using different functions, information will inevitably be destroyed in each instance of transmission. In this light, we may think of cultural attractors as reflecting a shared reference frame that allows cultural information to be preserved and potentially built upon over time.

Imagine a first individual who invents a dance, focusing primarily on their fancy footwork. An observer with a very different cognitive landscape may, frustratingly, fail to appreciate the first dancer's footwork at all, but instead attend to their arm movements, and therefore end up "recreating" a very different dance. A third individual may attend mainly to the second dancer's head movements, and so on. It is not that these individuals are copying inaccurately per se, but instead that they do not even agree on what it is they are supposed to copy. If we posit some cognitive function that transforms sensory signals into new behaviors—a cognitive landscape—these individuals have different, but equally valid, functions. In such a situation, there may be social learning occurring in some sense (or at least social influence), but variants would not be expected to cluster in any identifiable way, and indeed it would be hard even to say there *exists* any cultural variant to evolve (a point made also by Claidière and Sperber, 2007). Conversely, when individuals within a group have highly aligned cognitive landscapes, productions of a cultural variant can differ substantially in "surface" characteristics while nonetheless retaining the same culturally relevant core. Consider a participant in a Western population who is asked to draw a smiley face using a red pen on a notepad, a second to copy this image using spray paint on a wall, and a third to copy the second using Lego blocks. In this case, they will all likely recognize each product as instance of the same culturally shared category, despite variation in the medium. In most respects—except just those few culturally relevant ones—these could be seen as "low-fidelity" copies. However,

the cultural core of these productions is not *in* the productions themselves, but instead is an abstract mental category shared across the individuals. When cognitive landscapes are aligned in this way, cultural transmission can occur with sufficient fidelity for selection to act on cultural variants in a Darwinian fashion.

1.1.2. Mechanisms of convergent transformation: The importance of collective cognitive alignment through enculturation

From our perspective, the most crucial insight of CAT is the point that social transmission is "reconstructive," meaning that cultural variants are not simply copied but actively reproduced by individuals, influenced in the process by the memories, biases, and proclivities present in their minds (Claidière et al., 2014; Sperber, 1996; Scott-Phillips et al., 2018). As described by Sperber (1996), reconstructive transmission may produce convergent transformation patterns when there is a "convergence of [...] affective and cognitive processes [...] of many people towards some psychologically attractive type of views in the vast range of possible views." We refer to this convergence as "collective cognitive alignment." In cases where cognitive alignment depends upon enculturation through experience, it becomes possible that cognitive alignment may *fail* to be achieved either within or across generations. This motivates the need for computational models of cultural attraction such as ours that do not make the assumption of pre-existing attractor points and instead appeal to a culture-cognition feedback loop.

Collective cognitive alignment through enculturation is not the only possible mechanism of convergent transformation patterns. Shared ecological factors are likely to produce cultural attractors in some cases, ranging from norms of sharing in harsh, isolated climates (Gerkey, 2013) to color categories in environments dominated by correlated spectral patterns (Baronchelli, Gong, Puglisi, & Loreto, 2010). Some cultural attractors may be driven by exogenous motivational factors, such as an imperial edict that results in widespread adoption of a particular hairstyle, upon penalty of death (Morin, 2016). And some cultural attractors may be the result of relatively universal *cognitive* attractors driven by genetic features under strong selection, such as an evolved salience bias for direct eye-contact leading to an increase in viewer-oriented figures over time in a portraiture tradition (Morin, 2013). However, humans exhibit tremendous cultural variation that cannot be attributed merely to ecological factors. As evidence of this, we could point to any example of warring, neighboring tribes that distinguish themselves with different cultural markers, languages, customs, and beliefs (Smaldino, 2019). Nor can we attribute this variability to genetic differences between populations, given that there is known to be more genetic variation within human groups than between (Lewontin, 1972). Therefore, we propose that it is critical to explain how cultural attractors may form as a result of the culture-cognition feedback loop in the absence of a strong determination by innate biases or shared ecological factors.

1.1.3. The problem of collective cognitive alignment

The process of cognitively aligning to a cultural reference frame—that is, of acquiring a set of categories and cognitive biases specific to members of a cultural community—is

often discussed as a purely individual-level learning process (Ashby & Maddox, 2005; Kuhl, 2000; Toscano and McMurray, 2010). The cultural background that provides the fodder for learning is assumed, at least by many cognitive scientists, to be generally stable. Individuals may vary but will observe similar training data and ultimately develop similar cognitive landscapes. But cultural environments, and the shared categories associated with them, can change over generational or even intragenerational timescales. As such, cognitive alignment is an ongoing collective coordination problem, in addition to being an individual learning problem.

There exist several computational models of the emergence of category conventions in groups (Baronchelli, Gong, Puglisi, & Loreto, 2010; Ke, Minett, Au, & Wang, 2002; Kirby, 2001; Puglisi, Baronchelli, & Loreto, 2008; Reali, Chater, & Christiansen, 2018; Steels & Belpaeme, 2005; Skyrms, 2010); reviewed in Contreras Kallens, Dale, and Smaldino (2018). However, these models assume that agents come pre-equipped with a shared set of recognizable and producible cultural variants, such that social transmission has perfect fidelity. In some cases, shared, fixed sets of signal and meaning categories are explicitly pre-defined, as in Kirby's (2001) iterated learning model. Several models have considered the coordination of linguistic labels for perceptual categories (Baronchelli et al., 2010; Gong, Baronchelli, Puglisi, & Loreto, 2011; Puglisi et al., 2008; Steels & Belpaeme, 2005), allowing perceptual categories to be flexibly adjusted through experience and for new linguistic labels to be created. However, in these models, the signals (e.g., verbal labels) are still transmitted with perfect fidelity, implying a globally defined set of signal categories that are available to everyone—a world of Platonic word forms. Even models that represent the possibility of transmission errors (Nowak, Krakauer, & Dress, 1999; Nowak & Krakauer, 1999) treat errors as confusions of one signal category for another, which again presupposes that individuals share a set of signal categories. While this modeling literature has produced many important insights, it does not address cases in which signal categories may be plastic and differ across individuals.

One attempt that begins to address the culture-cognition feedback loop is a model of phonemic evolution by Winter and Wedel (2016). In their model, two agents each possessed a mental model of the set of phonemes in their language, represented as labeled clusters of two-dimensional (2D) point exemplars stored in memory. As the two agents communicated by producing signals to each other under the influence of cognitive biases, each agent categorized and stored new exemplars received from their neighbor while prior exemplars decayed in memory. In the process, the agents' labeled clusters of exemplars drifted around the signal space, corresponding to the co-evolution of individuals' perceptual distinctions along with a shared lexicon. While this model is a strong step towards giving due diligence to the issue of cognitive alignment in cultural evolution, Winter and Wedel's (2016) agents *begin* each simulation aligned, and therefore their results can tell us little about how cultural attractors initially emerge. Furthermore, with just two agents interacting in a highly constrained manner, their model cannot address how a cultural attractor landscape is generated and maintained within a dynamic population.

Populations in which cultural attractors emerge often involve non-static sets of individuals. Old members die or leave, while new members are born or arrive from elsewhere. Consider

that young learners, by definition, contribute different cognitive biases to the cultural attractor landscape than seasoned 'experts,' such that deaths and departures of the old and an influx of new learners threaten to alter a cultural attractor landscape in potentially drastic ways. If too many learners enter the population too fast, or many experts suddenly die, a cultural attractor landscape can change or even dissipate (unless there are other stabilizing factors, for example, mechanisms for external information storage). This is a central point of the 'linguistic niche hypothesis,' which holds that languages adapt to their learners, in addition to the reverse process (Bentz & Winter, 2014; Dale & Lupyan, 2012; M. L. Lewis & Frank, 2016; Winters, Kirby, & Smith, 2015). For example, it has been proposed that as linguistic populations expand, they may incorporate a greater proportion of adult learners, causing pressures for language to change as a result of different cognitive biases between adults and children (Dale & Lupyan, 2012; Reali et al., 2018). Thus, language (and culture more generally) should not be thought of as information passively transmitted from one generation to the next, but instead as a complex adaptive system, wherein variants are products of individual cognitive landscapes, and individual cognitive landscapes are shaped by experience with other variants (Enfield, 2014; Beckner et al., 2009).

In summary, we argue that understanding how cultural attractors can emerge and stabilize in the absence of innate cognitive attractors is an important step towards understanding the capacity for cumulative cultural evolution. Explaining complex processes requires mechanistic formalization (Epstein, 1999; Smaldino, 2017), but any initial formalization is likely to be incomplete, as models tend to accumulate nuance iteratively. Below, we present a model that we believe lays the groundwork for understanding the emergence of cultural attractors in the absence of strong determination from innate biases or shared ecologies. In a population of interacting, cooperative individuals within a cultural community, it is reasonable to assume that mutual understanding is often, if not always, the goal of communication. Individuals will develop categories based on what is communicated to them, and use those categories to communicate similar information to others. We have argued that the existence of shared cognitive biases is a prerequisite for treating cultural transmission as inheritance with noise, and so we do not appeal to selectionist principles in developing our theory. Instead, we model the intertwining of cognitive, communicative, developmental, and demographic dynamics. Because many mechanisms that allow for these dynamics are themselves evolved (e.g., learning periods and life cycles, social tendencies, neural structures), a full explanation must eventually reintroduce selection processes, but these we save for future work.

Our model currently offers only a general mechanism by which collective cognitive alignment may emerge through general principles of communication and learning and should not be taken as mapping precisely onto specific empirical patterns. In other words, ours is a "how possibly," rather than a "how actually," model (Craver, 2006). We see this model as complementary to the careful historical and anthropological work associated with CAT, which describes distinct instances of cultural attraction and identifies explanatory forces, and we suggest that our model may be extended in future work to formalize how specific perturbations or parameters noted in the CAT literature can influence a cultural attractor landscape.

2. Model description

Our model is intended to represent multiple generations in a population of individuals that interact and observe one another, implicitly shaping each other's cognition in the process. The basic requirements for modeling such a system include (1) a population of individuals, (2) a process whereby agents age and die, and new agents are born, (3) a mechanism for individuals to interact and observe one another, (4) a representation of the systems that shape individuals' perception and production of information (i.e., cognitive landscapes), and (5) a mechanism for updating these representations based on experience (i.e., learning).

We begin with a population of N agents, arranged as nodes in an undirected network where edges represent opportunities for communicative interactions. Four network structures were explored, with the default being fully connected (more details below). Each agent has an age, which is represented in our model as the number of time steps for which it has been 'alive.' All agents are initialized with an age of zero.

The model dynamics occurred in discrete time steps (illustrated in Fig. 2), each of which consisted of two stages: *communication* and *reproduction*. In the communication stage, we iterate through agents in order of their position on the network, giving each a turn to communicate a signal to a randomly selected neighbor. Each communicator randomly selects one neighbor for interaction (the receiver) and produces a signal, which may be distorted by noise. Receivers then learn something from the observed signal.

A perceptual signal is some pattern of activity across the sensory receptors of an observer. Patterns of activity across n sensory receptors can be represented as points in an n-dimensional space. While the number of sensory receptors may be large, we assume that we can obtain a projection of this space onto two dimensions for plotting, which is commonly done in connectionist models of cognition and neuroimaging work, using mathematical tools such as principal components analysis. Thus, we represent signals as real-valued points on a 2D $S \times S$ square (we used $S = 100$). These axes could correspond to any two featural dimensions which may be extracted by a category learning system, such as the voice-onset-time and fundamental frequency of a speech token (Toscano and McMurray, 2010) or the length and width of an arrow (Henrich et al., 2008).

Signal perception and production are both served by a learned representation of category structures. There are many ways to model category representation and learning. Here, we utilize an unsupervised, 2D mixture of Gaussians (MOG) model adapted from Toscano & McMurray (2010), which they found to effectively model the acquisition of phoneme categories in English. We expect this algorithm could be replaced by many cognitively plausible models of categorization, including exemplar-based models (e.g., Winter & Wedel, 2016) or neural network classifiers (e.g., Steels & Belpaeme, 2005), without changing the overall picture. However, a MOG has useful mathematical properties and can capture complex distributions with relatively few parameters, so it may be less computationally intensive than other models.

Each agent i possesses in memory a MOG of size $K = 20$, where each category k is defined as a 2D Gaussian distribution with a mean μ_{ik}, standard deviation σ_{ik} (both mean and standard deviation are 2D vectors), a correlation ρ_{ik} between dimensions (though for simplicity,

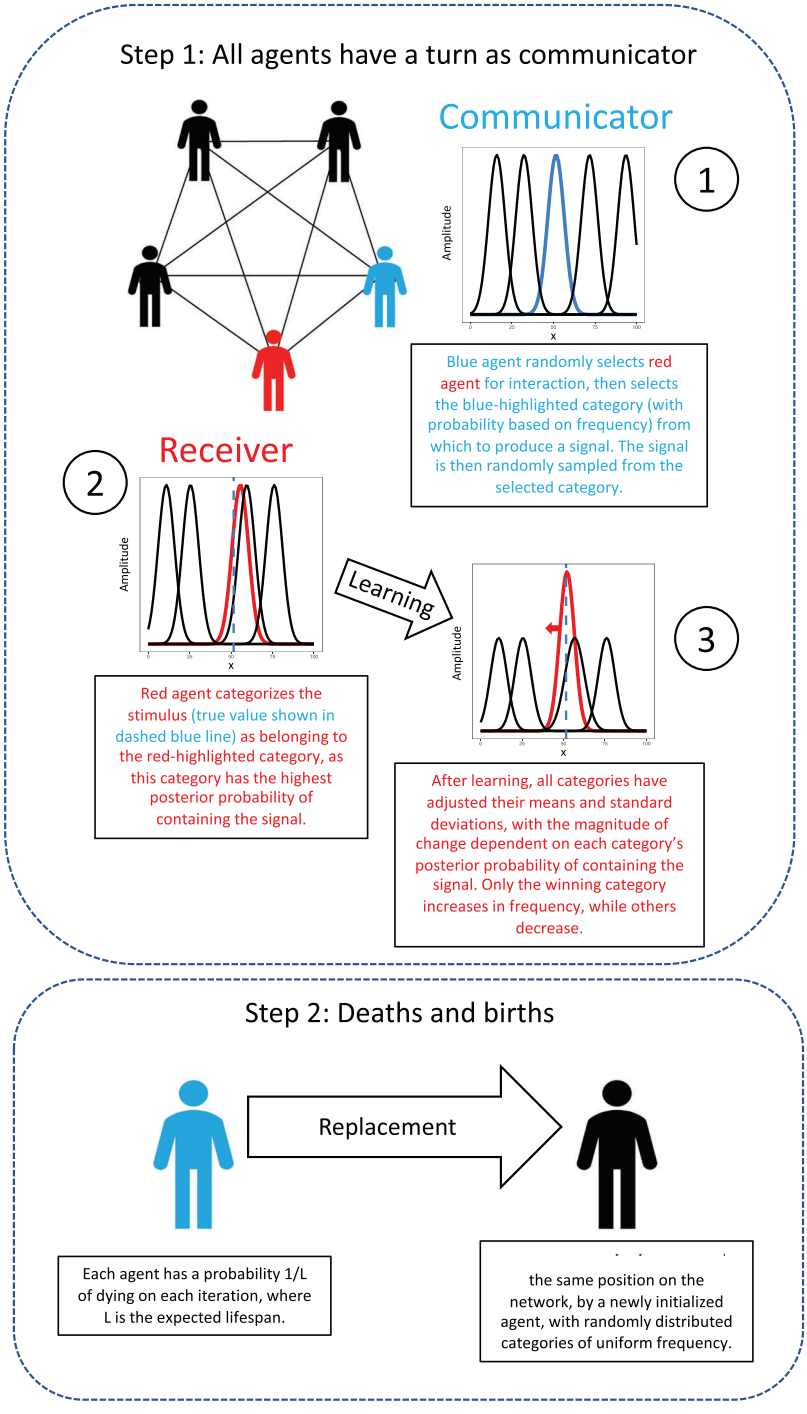

Fig. 2. An illustration of the model dynamics.

we chose to keep ρ fixed at 0), and an amplitude ϕ_{ik}. The mean of each Gaussian represents the central tendency of the category (similar to prototypes in some theories of categorization), while the standard deviation represents the variability of the category, with smaller standard deviations equating to more specific categories. The amplitude ϕ_{ik} represents the prior probability that a random stimulus is a member of that category. At initialization, the mean of each category for each agent is randomly drawn from a uniform distribution in [[0, 100][0,100]], with a fixed standard deviation of $\sigma_{\text{initial}} = 5$. The amplitude of each category is initialized at $1/K$ so that all categories are initially equally probable.

When acting as communicators, agents generate a signal by sampling a category from their MOG, with the probability of selecting each category proportional to the estimated prior probability of observing that category (the amplitude ϕ_{ik}; we describe below how this is estimated through observation). This assumes that agents simply attempt to reproduce the same frequency distribution that they have learned. This is a reasonable starting assumption for the many cultural domains in which imitation and conformity are useful, such as language, but other ways of mapping from memory to production should be explored in future work. We assume that communicators attempt to signal the mean of their selected category, but that noise may distort the signal that gets received. We used simple Gaussian noise added independently to each signal dimension, with a mean of 0 and a standard deviation of W.

Upon receiving a signal from a communicator, the receiver agent uses Bayesian inference to categorize the signal and adjust the parameters of their MOG representation in memory. This process is somewhat complicated and is described in greater detail below (see "Learning"). Essentially, the receiver first maps the signal as a member of the most likely of its own stored categories. It then updates the properties of its categories to reflect this new information.

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. Each agent has a probability of $1/L$ of dying at each time step, implying an expected lifespan of L time steps. Any agent who dies is removed from the simulation and replaced by a new agent. Newly born agents are initialized in the same way as agents at the beginning of each simulation.

Each simulation was run for 40,000 time steps. Based on piloting, this length appeared sufficient for most of our outcome measures to stabilize. The procedures used to analyze the model are described in detail in the Outcome Measures section. The code to run this model is available on OSF.[1]

2.1. Learning

Upon receiving a signal from a neighbor, receiver agents categorize the signal and adjust their category representations using Bayesian inference. Agents first compute the likelihood of the signal belonging to each category j in their MOG, according to a Gaussian likelihood

function G:

$$G_{ij}(x,y) = \phi_{ij}\left(\frac{1}{2\pi\sigma_{ijx}\sigma_{ijy}\sqrt{1-{\rho_{ij}}^2}}\exp\left(-\frac{1}{2\left(1-{\rho_{ij}}^2\right)}\left(\frac{\left(x-\mu_{ijx}\right)^2}{{\sigma_{ijx}}^2}-\frac{2\rho_{ij}xy}{\sigma_{ijx}\sigma_{ijy}}+\frac{\left(y-\mu_{ijy}\right)^2}{{\sigma_{ijy}}^2}\right)\right)\right). \tag{1}$$

The likelihood of each category can be thought of as the goodness-of-fit of the signal to each category in the agent's repertoire. In neural network terminology, we can think of the likelihoods as the activation levels of each output node (each category) in response to the input signal. The marginal likelihood M of the signal is the sum of the likelihoods over all categories in an agent's MOG (or we can think of it as the sum activation at the output layer of a neural network):

$$M_i(x,y) = \sum_{j=1}^{K} G_{ij}(x,y). \tag{2}$$

And the posterior probability P of each category is then calculated as the ratio of the likelihood to the marginal likelihood:

$$P_{ij}(x,y) = \frac{G_{ij}(x,\ y)}{\sum_{j=1}^{K} G_{ij}(x,y)}. \tag{3}$$

The posterior probability is the proportional goodness-of-fit of the signal to each category or the activation of each category scaled by the total activation across all categories. The category with the highest posterior probability (the 'argmax') can be thought of as the label an agent applies to a signal or their 'interpretation' of a signal.

The parameters of all categories are then updated using a gradient descent algorithm. This algorithm acts to maximize the marginal likelihood function M by adjusting parameters along the derivative of M with respect to each parameter. More simply stated, agents move their categories around in the 2D signal space and adjust their shapes such that the signal would be better fit by their MOG, if the agent received the same signal again. Importantly, the magnitude of the adjustment on each category is scaled by its posterior probability. This means only categories that are probable given a signal are moved, while others change little, which prevents all categories from converging to a single point. The learning rules for each parameter are as follows:

$$\triangle\mu_{ijx} = \eta_\mu P_{ij}\frac{1}{\left(1-{\rho_{ij}}^2\right)}\left(\frac{x_{ij}-\mu_{ijx}}{\sigma_{ijx}^2}-\frac{\rho_{ij}y_{ij}}{\sigma_{ijx}\sigma_{ijy}}\right), \tag{4}$$

$$\triangle\sigma_{ijx} = \eta_\sigma P_{ij}\left(\frac{\left(x_{ij}-\mu_{ijx}\right)^2}{\sigma_{ijx}^3\left(1-{\rho_{ij}}^2\right)}-\frac{\rho_{ij}\left(x_{ij}-\mu_{ijx}\right)\left(y_{ij}-\mu_{ijy}\right)}{\sigma_{ijx}^2\sigma_{ijy}\left(1-{\rho_{ij}}^2\right)}-\frac{1}{\sigma_{ijx}}\right), \tag{5}$$

where η represents the learning rate for each parameter. For added simplicity in visualization and the signal production process, correlations between the two dimensions of each category were fixed at 0 and did not update.

Unlike the means and standard deviations, the amplitude (also equivalent to a Bayesian prior) parameter ϕ was updated based on winner-takes-all competition, such that only the category with the highest posterior probability increased in amplitude. Intuitively, this means that agents treat each signal as having actually come from only one category, such that each observation should only increase the estimated base rate of one category. The amplitude of the winning category is updated according to the following learning rule:

$$\triangle\phi_{ij} = \eta_{\phi} P_{ij}(x, y). \tag{6}$$

After updating the amplitude of the winning category, the amplitudes across all categories were normalized. This winner-takes-all competition increases the amplitude of frequently heard categories while suppressing unused categories. McMurray, Aslin, and Toscano (2009) showed that this type of competition is crucial for unsupervised learning when the number of categories is unknown; in the absence of winner-takes-all competition, individual learners were unable to detect the correct number of phonetic categories within their training data. As a whole, these learning rules allow agents to begin with a relatively large number of equally probable categories (e.g., 20), and over time to pare their category representation down into the simplest structure that effectively captures the distribution of signals they observe.

We also explored the effects of a 'critical period' in learning. The critical period refers to a period early in life during which the brain is highly plastic and learning is facilitated. The existence of such a period is well established in the literature on language development and may be an important factor in cumulative culture. This was implemented by turning off learning for an agent after they reached an age C in time steps.

It should be noted that, while our agents use Bayesian inference to categorize signals, we take this to be an algorithmic-level description of cognitive operations, in line with arguments presented by McClelland et al. (2010). The mechanism(s) underlying these inferences could be implemented by a distributed neural network or other system, and hence we need not take a stance with respect to the cognitive reality of Bayesian inference here.

2.2. Network structure

We explored four different network structures, illustrated in Fig. 3, to examine the ways that connectivity patterns can influence cultural attractor dynamics. All networks were undirected, meaning that links were bidirectional, and network structure was held constant throughout each run. In our baseline model, agents were arranged in a fully connected network. This results in the largest possible mean degree of $N - 1$ (here, 49), and the largest possible clustering coefficient—the average proportion of agent i's neighbors who are also connected to each other—of 1. This fully connected network also has the smallest possible an average shortest path-length—the average of the minimum edges traversed to connect any two nodes—of 1.

We next considered a connected caveman graph (Watts, 1999), in which agents were first arranged into five fully connected 'cliques' of 10 agents each (meaning each agent has nine neighbors). In each clique, one edge is randomly rewired to a neighboring clique, such that the cliques are ultimately connected in a loop. This network has a near-maximal clustering

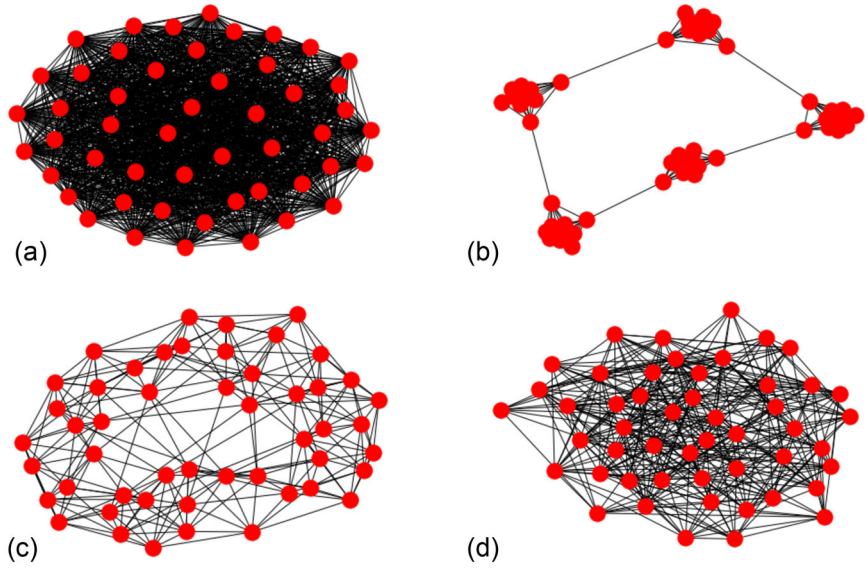

Fig. 3. Four network structures explored in our model: (a) a fully connected network; (b) a connected caveman network; (c) a small-world network; (d) a realistic social network.

coefficient of .936. The average shortest path length, however, becomes much longer than the fully connected network, reaching 3.37.

The third network explored was a small-world network (Watts, 1999), which is formed by connecting each agent to their nearest N neighbors, then randomly rewiring connections with probability P. We used a network where each agent had 10 neighbors and the rewiring probability was set to .1. This resulted in networks with, on the average of 1000 samples, a clustering coefficient of .51 and average shortest-path length of 2.18. All agents had 10 neighbors.

Finally, following methods used by Reali et al. (2018), we explored a 'realistic' social network. These networks had a connectivity pattern inspired by empirical patterns seen in modern populations, which have indicated that average nodal degree (i.e., average number of neighbors individuals have) scales with population size such that the clustering coefficient is invariant at a value of $\sim$ .25 (Schläpfer et al., 2014). We constructed 20 such networks, which were sampled from randomly across the 100 runs of the model. These networks had a mean clustering coefficient of .261 and a mean average shortest path-length of 1.7. Agents had an average of 15 neighbors.

2.3. Outcome measures

We analyzed the emergent cultural attractor landscapes along three dimensions: (1) cultural complexity, which we operationalize simply as the number of categories detected at the population level; (2) cultural stability, or the rate of change of the category distribution in signal space; and (3) cognitive alignment, meaning the similarity of cognitive landscapes across

individuals. These measures were chosen because of their applicability to a wide array of phenomena in cultural evolution research. First, cultural complexity may relate to the combinatorial possibilities of a cultural repertoire, and measures of complexity are often appealed to in discussions of cumulative cultural evolution. Second, stability may be important for the accumulation of new cultural variants that depend upon existing ones, or for the possibility of intergenerational transfer of information (e.g., if a language changes drastically every generation, communication between individuals of different generations may be disrupted). Third, cognitive alignment may be related to the degree of specialization versus generalization of knowledge in a community, and different domains may benefit from different degrees of alignment (e.g., language is most useful when it is widely shared, while engineering feats may benefit from the joint efforts of individuals with different knowledge).

To obtain our measures, the model was observed every 1,000 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. One hundred runs were conducted for each parameter setting. 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 to 50. We then used the gap statistic (Tibshirani, Walther, & Hastie, 2001) to select the optimal value of k at each timepoint. The optimal value of k was used as an estimate of the complexity of the cultural attractor landscapes.

Next, to examine the stability of the cultural attractor landscape, 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. In this way, the EMD is a type of optimal transport algorithm. While there are many popular similarity metrics to choose from, such as the Kullback–Leibler (KL) divergence or Jensen–Shannon (JS) divergence, we selected the EMD because it is symmetrical (unlike KL) and can handle events with a probability of 0 (unlike JS). Furthermore, the EMD accounts for the metric space in computing distances. For example, two distributions of the same shape but located in different regions of the signal space will be treated as different under the EMD but would have a distance of 0 under KL divergence, because the latter does not account for the location of the observations.

Because our signal space is continuous, to compute the EMD 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 20×20 evenly spaced points (each square being 5×5) and the number of observations in each square was counted, creating a 2D histogram which was then normalized to sum to 1. 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 1,000 timesteps).

Finally, to examine the cognitive alignment across agents, we computed the average EMD of the distribution of signals generated by an individual agent to the distribution generated by the rest of the population. Since this is a dissimilarity metric, we will henceforth refer to this measure as cognitive *dis*alignment. At each evaluated timepoint, a 2D histogram was constructed from the signal samples from each individual agent i in a population of size N and was compared to another histogram constructed from the signal samples corresponding to every agent *besides* the focal agent (similar to the 'jackknife' resampling technique). Finally, we took the average of these values across agents, which provides a measure of the relative cognitive alignment versus idiosyncrasy or generalization versus specialization, in a population.

3. Simulation experiments

In this section, we first present a qualitative analysis of the model dynamics. We then consider three case studies illustrating applications of the model to several areas of inquiry within cultural evolution. First, we consider the effect of transmission noise, which we find has the effect of stabilizing cultural attractor landscapes. Then, we consider the effect of longer lifespans and critical periods in learning and find that shorter learning times may generate more complex and more stable cultural attractor landscapes. Finally, we consider the effect of population size and network structure. We find that large populations stabilize and simplify cultural attractor landscapes, while highly cliquish network structures can allow the maintenance of many distinct cultural categories.

3.1. Baseline model: Qualitative analysis and visualization

In order to get an intuitive sense of the dynamics of our model and how the emergent patterns act as cultural attractors, we will first visually analyze the behavior of the model over time on a single representative run (see Tables 2 and 3 for parameters). Figure 4 shows the state of all categories across all agents at nine different time points during a single representative run.

The model begins with all agents possessing a set of equal-amplitude categories, uniformly distributed throughout the signal space. Over the first 5,000 time steps, we can see cultural attractors beginning to emerge, as nearby categories are pulled closer and competition at the cognitive-level results in some categories getting suppressed, while others increase in amplitude (and, therefore, the probability that they will be produced in the future). By 10,000 generations, a clearly distinguishable set of tight clusters have emerged, though there remain some looser clouds of low-amplitude categories, likely driven by new learners entering the population (see Fig. 8). At this point, the model appears to have reached a dynamical equilibrium, where the qualitative pattern remains the same, but clusters continue to drift around stochastically. Some categories move too near to each other and "merge," while new clusters may occasionally arise in empty regions and others occasionally fade away. 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 increas-

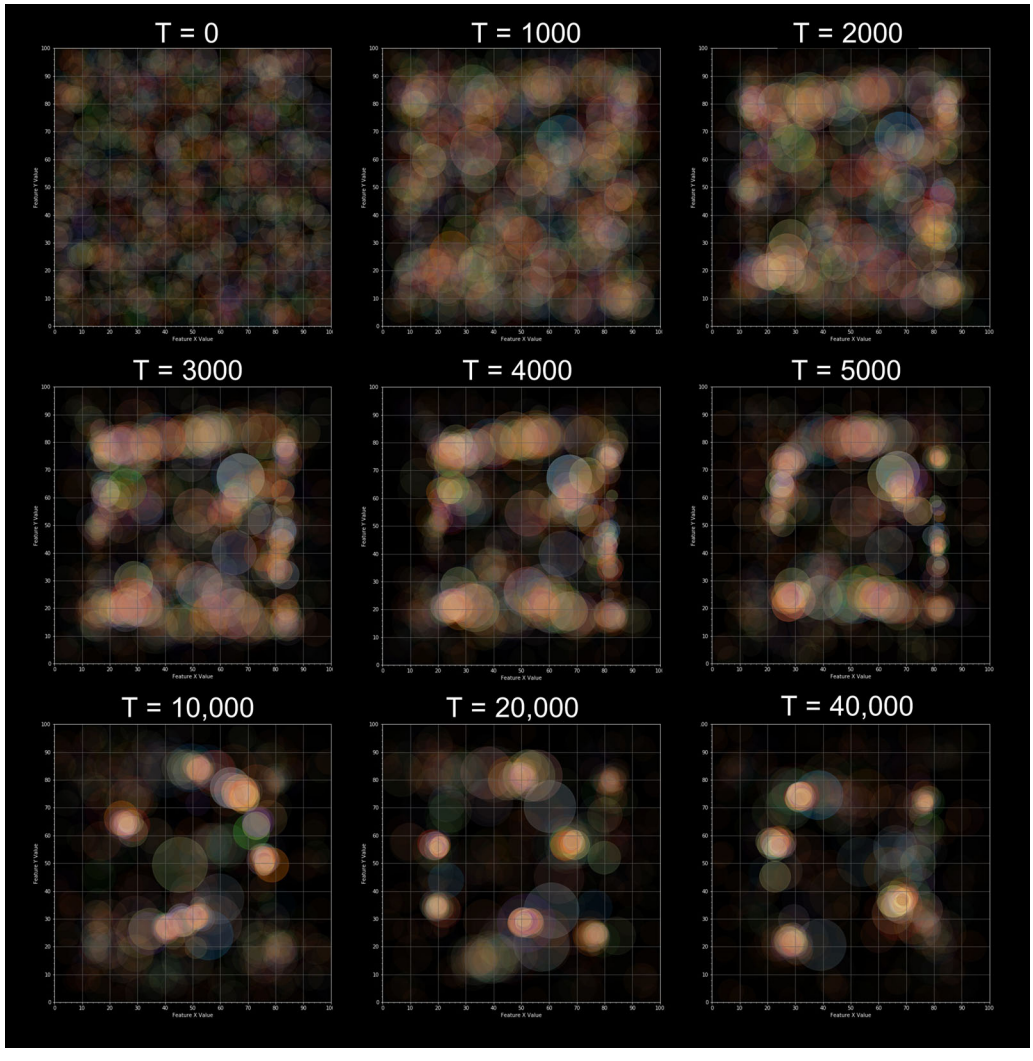

Fig. 4. The states of all categories across all agents in a population at nine time points of one run. Different colors correspond to different agents (here $N = 50$ agents). Each agent has multiple categories (here $K = 20$ categories) in their MOG, which are individual points ($20 \times 50 = 1{,}000$ points total). The size of points is proportional to the *SD* of the category, and the transparency (alpha value) of the points is proportional to the amplitude of the category, such that low-frequency categories become more transparent. The appearance of fewer points in later time steps is the result of both of the alignment of categories across agents, such that points overlap, as well as the fact that most categories in each agent's MOG become suppressed, rendering them transparent in the plot. It should be noted that, because both overlap and amplitude impact the transparency of points, their respective contributions cannot be visually distinguished (i.e., the same visual result can be achieved by fewer overlapping points of greater amplitude or more overlapping points of lesser amplitude).

Table 2
Variable model parameters. The values used in the baseline model are presented in bold font

Parameter	Values Explored	Description
W	**0**, 3, 10	*S.D. of Gaussian noise*. In transmission of a signal, Gaussian noise is added with mean 0 and S.D. $= W$.
L	5,000, **10,000**, 15,000	*Expected lifespan*. On each iteration, each agent has a probability of $1/L$ of 'dying' and being replaced by a new agent.
C	2,500, 5,000, 10,000, 20,000, **40,000 (length of simulation)**	*Length of the 'critical period.'* After reaching age C (in time steps), learning is turned off for an agent.
N	10, 25, **50**, 100, 200	*Population size*.
Network type	**Fully connected**, connected caveman, small world, realistic social network	Four different network structures were explored for connecting agents to neighbors in communication.

Table 3
Fixed model parameters

Parameter	Value	Description
K	20	Number of categories in each agent's MOG
σ_{initial}	5	The S.D. of each category in an agent's MOG upon initialization
η_μ	1	Learning rate for category means
η_σ	1	Learning rate for category standard deviations
η_ϕ	0.001	Learning rate for category amplitudes
η_ρ	0	Learning rate for correlation between dimensions. For simplicity, this value was held constant at 0 such that the two dimensions were uncorrelated.

ing in amplitude 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 encompasses the former two. (See the Supporting Information for a video version of Fig. 4.)

This dynamical equilibrium is made clear when visualizing the number of clusters that are detected at the population level over time. Fig. 5a shows a time series of the raw number of clusters detected using the k-means algorithm and the gap statistic (with a maximum k of 50), averaged over 100 runs with the baseline parameter settings. This plot shows that our cluster detection algorithm settles at $\sim$ 15 clusters by 20,000 time steps. Fig. 5c reveals that cognitive disalignment also stabilizes within approximately the same time frame. However, Fig. 5b shows that the distribution of categories throughout the signal space continues to change at roughly the same rate over the entirety of each run. Given that the number of categories detected and the average disalignment of agents appear to reach equilibrium by

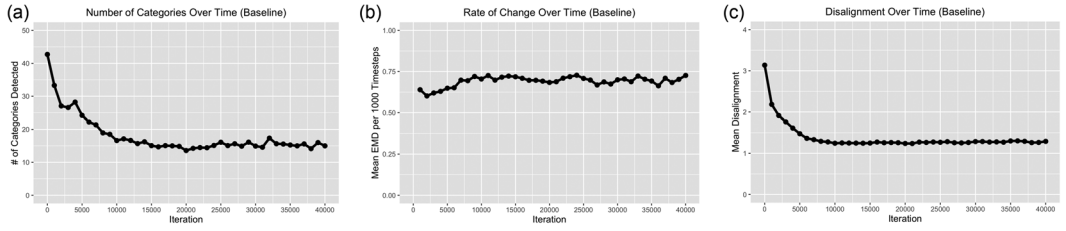

Fig. 5. Time series, with each point representing the average over 100 runs of the baseline model of (a). The complexity of the cultural attractor landscape. (b) The rate of change of the cultural attractor landscape over time. (c) The cognitive disalignment of agents to the population distribution.

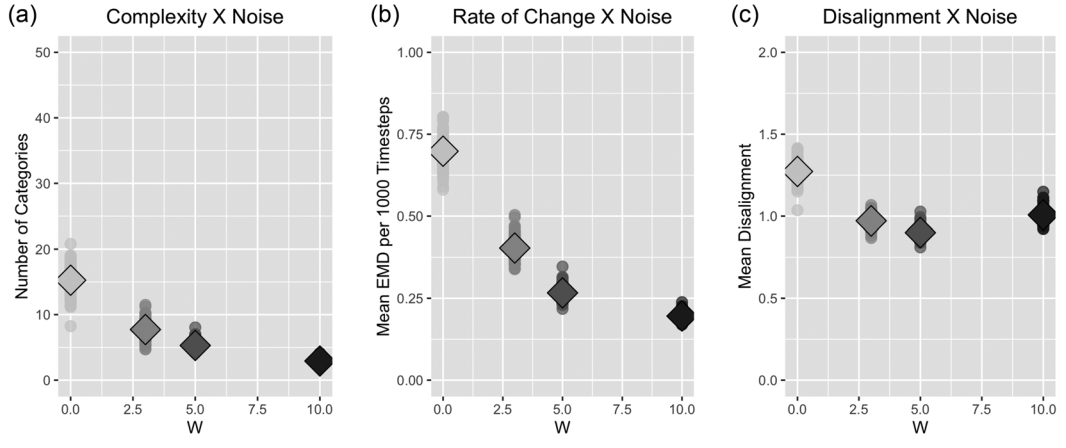

Fig. 6. The effect of variable Gaussian transmission noise with mean $= 0$ and SD $= W$ on (a). The complexity of the cultural attractor landscape. (b) The rate of change of the cultural attractor landscape over time. (c) The cognitive disalignment of agents to the population distribution.

20,000 timesteps, all subsequent analyses used average values over the final 20,000 timesteps (the second half) of any given run.

We can think of the clusters that form in our model as cultural attractors because these global patterns are precisely what individuals learn to approximate, and thus the clusters are attractor points in cognitive development. Of course, as others have already stated, these cultural attractors are simply statistical aggregates; individual agents do not have direct access to the population-level attractors but only to unique signals. However, because these clusters correspond to the expected distribution of observations for a random agent (in a fully connected network), these statistical abstractions constitute a real force shaping cognition.

3.2. Some noise is beneficial for stabilizing cultural attractor landscapes

We find that as transmission noise is increased, the cultural attractor landscape becomes increasingly stable (Fig. 6b) This effect is due to the fact that, as noise increases, agents less reliably signal the true mean of their categories, which slows the rate of learning and therefore

the rate of change at the global level. Understandably, increasing noise is also associated with a decrease in the complexity of the cultural attractor landscapes, because when categories become more diffuse, fewer of them can be maintained in the same space (Fig. 6a). Some noise (e.g., $W = 5$) also helps to facilitate cognitive alignment in the population (Fig. 6c), because the slower moving targets for learning make it easier for agents to acquire all of the categories in their population. However, the effect of noise on alignment is non-linear: we observe a slight *increase* in disalignment when noise is increased from $W = 5$ to $W = 10$. This suggests that, when $W = 5$, the complexity of the cultural attractor landscape is sufficiently low, and the rate of change sufficiently slow, that agents can effectively align to the population pattern, and therefore further increases in noise will merely reduce the complexity of cultural attractor landscapes at no additional benefit.

3.2.1. Discussion

Research on cultural evolution often focuses on the issue of transmission fidelity: transmission noise is generally considered to be a limiting factor for the purposes of cumulative cultural evolution (Nowak, Krakauer, & Dress, 1999), and the success of human populations in developing complex cultural repertoires is often attributed to the high fidelity with which we can transmit information, relative to other species (H. M. Lewis & Laland, 2012). At the same time, many fields outside of cultural evolution have seen a growing recognition of the crucial role that noise can play in complex dynamical systems. This point is exemplified in the literature on 'stochastic resonance,' which emphasizes that some amount of noise is beneficial for the detection of weak signals in non-linear systems (Gammaitoni, Neri, & Vocca, 2010; McDonnell & Ward, 2011; Wiesenfeld & Moss, 1995). For example, work by Goldman (2004) has shown that the possibility of synaptic transmission failures in the brain can actually enhance the informational efficiency of a synapse.

The behavior of our model with respect to noise suggests a bridge between the literature on transmission fidelity and the work on stochastic resonance. We find that as transmission noise is increased, and the cultural attractor landscape becomes increasingly stable over time. These effects are due to the fact that, as noise increases, agents signal the true mean of their categories less reliably, which slows the rate of learning, and therefore the rate of change at the global level. In turn, this helps to promote cognitive alignment across individuals, because the global pattern becomes a slower-moving target for learning. In a domain such as a language, cognitive alignment is of crucial importance, and thus it appears that transmission noise may play a role in the self-organization of linguistic conventions (at least at the level of speech sounds). If there is too *little* noise in transmission, categories may change so rapidly as to create problems using these categories in higher order systems. For example, lexical categories, which are signaled by combinations of phonemes, may not be possible if phoneme representations are highly unstable in a population.

However, our results should not be taken as contradictory to research suggesting that transmission fidelity is the "key to the build-up of cumulative culture" (H. M. Lewis & Laland, 2012). Rather, we suggest that moderate amounts of noise at the level of behavior/perception promote stable categories that are broadly shared, which counterintuitively makes these categories able to be signaled with enhanced fidelity. In other words, some within-category noise

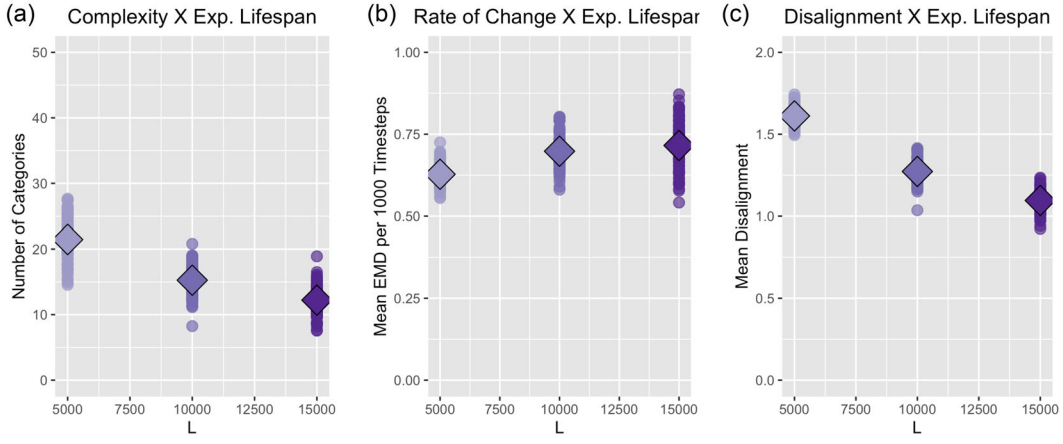

Fig. 7. The effect of variable lifespans L on (a) the complexity of the cultural attractor landscape, (b) the rate of change of the cultural attractor landscape over time, and (c) the cognitive disalignment of agents to the population distribution.

allows categories to become more distinguishable overall. It is important to reiterate that the cultural attractor landscapes in our model are akin to perceptual distinctions and should not be confused with higher order cultural variants that are transmitted *by virtue of* shared perceptual categories. As such, our results suggest that some noise at the level of perception/production may be important for ensuring transmission fidelity at higher levels of abstraction.

3.3. Longer learning times can result in decreased complexity of cultural attractor landscapes, and critical periods can enhance their stability

While longer learning times intuitively seem necessary in order to acquire more complex knowledge structures, somewhat surprisingly, we find that the complexity of cultural attractor landscapes *decreases* as learning times grow longer. We can see this effect in Fig. 7a, where expected lifespans varied and learning proceeded over the full lifespan. This effect is due to the fact that longer learning times also allow more time for cognitive competition between categories to proceed, which results in more categories becoming suppressed. This suggests that, as agents grow older, they eventually underfit the population distribution, possessing only a subset of the categories that are active at the population level (this is reflected in Fig. 8, which shows that older agents conform more poorly to the population distribution than middle-aged agents). Over generations, as new learners are influenced by the behaviors of their older neighbors, this results in a continued decline in the number of categories that are present. However, this comes with the potential benefit of promoting cognitive alignment overall, as agents can more readily fit the global distribution when it is simpler (Fig. 7c).

We next considered the effect of adding a critical period of learning, which was implemented by turning off learning after an agent passed C iterations in age. We find that critical periods moderate a trade-off between complexity (Fig. 9a) and stability (Fig. 9b) of the

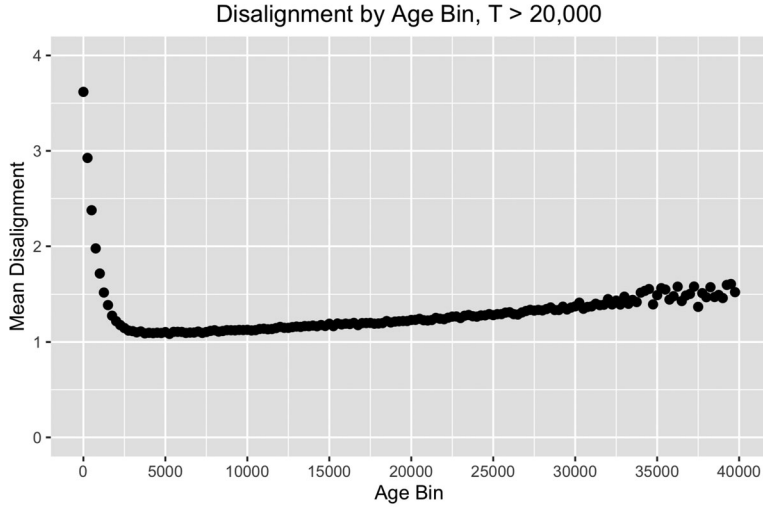

Fig. 8. Cognitive disalignment with respect to the population clustering pattern by age.

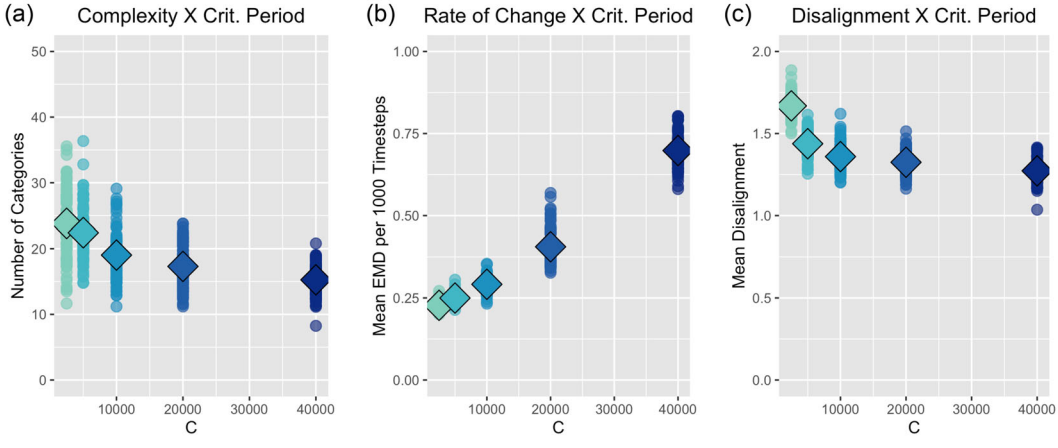

Fig. 9. As a function of the length of the critical period, *C*: (a) The complexity of the cultural attractor landscape, (b) the rate of change of the cultural attractor landscape over time, and (c) the disalignment of agents to the population distribution.

cultural attractor landscapes. This occurs because, when learning is restricted to a subset of the lifespan, agents who have stopped learning can remain in the population to act as stable models for more recently introduced learners. Fig. 9b shows that the equilibrium value of the rate of change increases as critical periods lengthen. However, if learning times are too short (e.g., 2,500 time steps in our model), we observe that agents do not have sufficient time to fit the population distribution, resulting in an increase in cognitive disalignment relative to moderate lengths of critical periods (Fig. 9c).

3.3.1. Discussion

The typical story told about learning in the research on biological evolution goes something like this: Investment in learning is helpful for adaptation to harsh and/or variable environmental conditions, but time spent learning is costly and detracts from reproductive opportunities. Thus, many organisms exhibit a "sensitive" or "critical" period early in life in which to assess environmental conditions, before committing to an adult phenotype (Frankenhuis & Panchanathan, 2011; Frankenhuis & Walasek, 2020; Panchanathan & Frankenhuis, 2016). In the domain of language, the existence of a critical period is one reason that language acquisition is facilitated in children and harder for adults (Birdsong, 1999; Hakuta, Bialystok, & Wiley, 2003). Such critical periods are generally thought of as a constraint, rather than an adaptation (Hurford, 1991; Komarova & Nowak, 2001). Our findings add complexity to this story, by revealing that as learning times grow longer (e.g., as an adaptation to a complex cultural repertoire), this may cause cultural attractor landscapes to simplify over time. If such a mechanism exists in real groups of cognitive agents, this could help to prevent runaway complexity: if cultural repertoires become complex, this may select for greater investment in learning, which may in turn result in the cultural repertoire simplifying. Thus, our findings suggest that shorter learning times may not only be selected for due to the cost of learning but also (likely at the group level) due to a possible role in stabilizing the cultural attractor landscape.

While it is possible that the effect of longer learning times on reducing the number of categories is merely an artifact of our learning algorithm and may not generalize to real human cognition, there is some reason to think that this may be a real effect. First, we can note that if learning in the brain is Hebbian, neuronal responses that have occurred in the past will increase the tendency for the same response to occur in the future, even if that response is inappropriate with respect to the input (i.e., a categorization error). For example, Japanese speakers may have trouble learning the contrast between /r/ and /l/ phonemes that are present in English, but absent in Japanese, because the presentation of either phoneme may simply reinforce the Japanese category that falls somewhere between the English /r/ and /l/ (McClelland, Thomas, McCandliss, & Fiez, 1999). In our model, when a stimulus is categorized as belonging to a high-prior-probability category that is slightly further away from the input value than a lower-prior-probability category, we may consider this a categorization "error," but the winning category will be reinforced nonetheless. Similar effects have been observed in humans, whereby making repeated responses that are in error results in a decrease in participants' abilities to discriminate between perceptual categories (McClelland et al., 1999). This point is further supported by evidence that older adults place a greater weight on lexical frequency when identifying a spoken word among a set of candidate words (i.e., older adults are more likely to identify the spoken word as corresponding to the more-frequent candidate; Revill and Spieler, 2012). Finally, we can note that aging is associated with a decrease in neural resources, which could further limit the number of perceptual distinctions available to an individual (Fjell & Walhovd, 2010). As such, the empirical research on aging and cognitive function suggests that the behavior of our model—a decrease in the complexity of the cultural attractor landscape as learning times increase—is plausible.

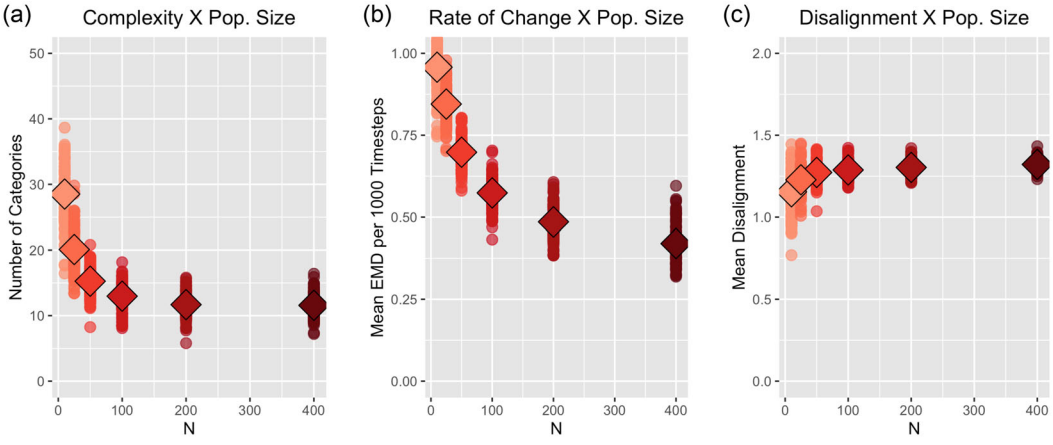

Fig. 10. The effect of variable population size, N on (a) the complexity of the cultural attractor landscape, (b) the rate of change of the cultural attractor landscape over time, and (c) the cognitive disalignment of agents to the population distribution.

Taken together, our results point to interesting trade-offs among stability, complexity, and cognitive alignment with respect to learning times and lifespans. When the problem space is continually changing, longer learning times may be beneficial. At the same time, longer learning times decrease the stability of the learning space but also may decrease the number of categories to be learned. Critical periods, on the other hand, may not only provide a fitness benefit by minimizing the energy invested in learning but may also play an important role in the stability and coherence of the cultural variants found in a population. It remains unclear how human developmental trajectories have evolved to balance these complex interactions in a way that allows for cumulative culture, but future explorations with our model, with the addition of representations of fitness and reproduction (allowing for heredity in cognitive capacities), may be able to shed some light on this issue.

3.4. Larger populations have simpler, more stable cultural attractor landscapes, and network structure can moderate these effects

Population size and/or density are commonly implicated as important factors in the potential for cumulative cultural evolution, with larger/denser populations being thought to sustain more complex cultural repertoires (Henrich, 2004; Reali et al., 2018). However, in our model we find that larger populations do not tend towards more complex clustering schemes (Fig. 10a).In fact, the pattern is quite the reverse, though the number of categories appears to approach a lower asymptote of ~ 10 categories as populations become large. Interestingly, the decrease in complexity that is associated with larger population sizes is not paired with a corresponding decrease in cognitive disalignment (Fig. 10c). Instead, disalignment shows a slight *positive* relationship with population size. This can be explained by the fact that, although larger populations appear to have simpler emergent category structures, agents

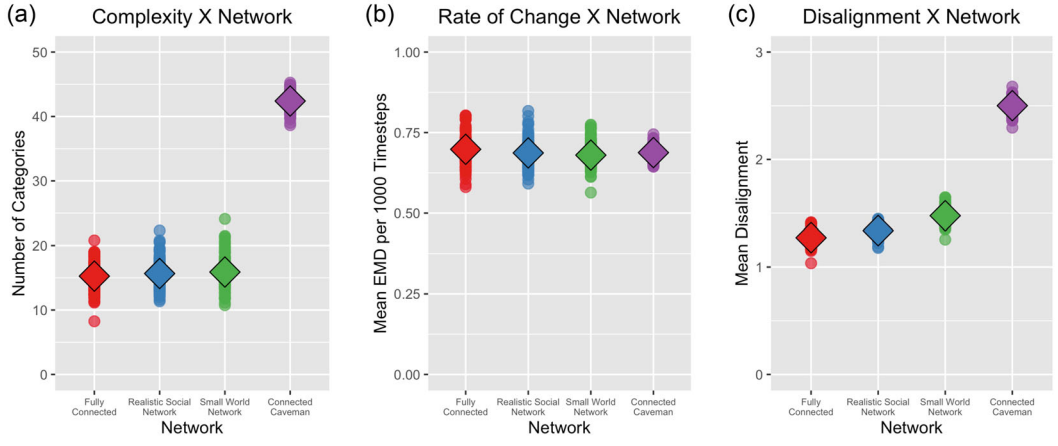

Fig. 11. As a function of the network type: (a) the complexity of the cultural attractor landscape, (b) the rate of change of the cultural attractor landscape over time, and (c) the cognitive disalignment of agents to the population distribution.

in larger populations also have fewer repeat interactions, which is a detriment to cognitive alignment. We also observe that larger populations have slower changing cultural attractor landscapes (Fig. 10b). This is because, in smaller populations, individual agents contribute more substantially to the global average. As such, deaths and births of new agents constitute a more significant perturbation in smaller populations, leading to sudden spikes in the rate of change.

Considering network structure, our results show no difference between fully connected, small-world, or realistic social networks in terms of the complexity of the cultural attractor landscape that forms (Fig. 11a). However, the connected caveman network differs dramatically from the others, showing far greater complexity. This effect is due to the fact that the limited connectivity between cliques in the connected caveman network limits the diffusion of conventions, such that each clique tends to converge upon a distinct set of categories, resulting in a much larger number of categories being maintained in the population overall. However, individuals *within* a clique do not actually have more complex cognitive landscapes, relative to individuals embedded in other networks. Thus, this effect is actually due to a decrease in the cognitive alignment of individuals with respect to the global pattern (Fig. 10c): individuals within a clique conform to each other, but not to others outside of their clique.

3.4.1. Discussion

Population size and demographic structure are some of the most commonly implicated factors in theories of cultural and linguistic evolution. For example, relating to population size/density, it has been proposed that larger and/or denser populations may be able to sustain more complex skills and technologies (Henrich, 2004) and that larger populations tend to develop larger vocabularies, but simpler grammars (Lupyan & Dale, 2010; Reali et al., 2018). Relating to network structure, the literature on group problem solving suggests that different

patterns of connectivity and/or network change are optimal for different types of problems (Lazer & Friedman, 2007; Rulke & Galaskiewicz, 2000; Smart, Huynh, Braines, & Shadbolt, 2010). For example, Lazer and Friedman (2007) showed that, in complex problem spaces where individuals can either independently explore the solution space or copy the solutions of successful neighbors, moderate amounts of network connectivity prove most efficient, because they balance breadth of exploration with the rapid diffusion of "good enough" solutions. Complementing this work, Smolla and Akçay (2019) recently showed that networks and culture may coevolve, with environments that select for specialist knowledge resulting in sparse connectivity patterns, such that individuals are repeatedly exposed to the same information and increase their depth of expertise, while environments that select for generalist knowledge result in dense connectivity patterns, for complementary reasons. Other recent results from Cantor et al. (2021) suggest that the relationship between network structure, population size, and diffusion mechanisms is highly complex: networks that perform best in terms of cumulative cultural evolution in one context may perform worst in another.

Some of the results of our case studies are consistent with existing research. For example, as in the work on group problem solving and the role of network structure on cumulative cultural evolution, we find that cliquish networks (like the connected caveman network) limit the diffusion of conventions, resulting in a greater diversity of cognitive landscapes in the population (Derex & Boyd, 2016; Lazer and Friedman, 2007). Other results are complementary to existing research. For example, to the best of our knowledge, there is no model that accounts for the fact that larger populations do not tend to have larger repertoires of perceptual categories (e.g., phoneme inventories: Creanza et al., 2015; Moran, McCloy, and Wright, 2012; though there is some debate here, for example, Fenk-Oczlon and Pilz, 2021), while they clearly do differ in the complexity of higher order cultural repertoires that depend on combinations of these categories, such as tools and grammar. While we have not currently explored the possibility of agents constructing artifacts that consist of combinations of elements, our model can be extended to allow for this possibility, as we will discuss in the next section. As such, our model can be integrated with existing models of cultural innovation, and therefore can allow for exploration of the interactions between these two levels of analysis. Finally, our model also produces some behaviors that have not been noted at all in the literature. For example, the role of population size on stabilizing change in cultural attractor landscapes, and the relationship between population size and cognitive alignment, are novel effects, to the best of our knowledge. Thus, our model may suggest interesting new paths for future research, in addition to complementing existing work.

Our explorations with network structure reveal how distinct patterns of cognitive alignment can arise from distinct patterns of connectivity. A network of connectivity determines not only how information flows through a population (i.e., the paths it takes through the network) but also how it is distorted as it flows through that network. Our model focuses on the patterns of distortion, but it is important to note that networks themselves may evolve, reaching different distributions of cognitive landscapes depending upon selection pressures in different domains. This could be a result of preferentially forming connections with those who are cognitively similar (e.g., because interactions are more successful on average) or selectively attending to

prestigious or knowledgeable others. Furthermore, networks of interaction for real individuals are better described as multiplex networks.

4. Conclusion

Despite some significant debates in the history of cultural evolution research, it is now generally agreed upon that both preservative dynamics (i.e., Darwinian selection) and transformative dynamics (i.e., cultural attraction) are crucial aspects of how culture evolves. We agree with previous claims by Henrich et al. (2008) and Claidière et al. (2014), that cultural attraction effects support Darwinian cultural evolution: when cultural variants cluster around points in the space of possible features, cultural information can be transmitted repeatedly without accumulating random error. Thus, while cumulative cultural evolution may depend upon high-fidelity *transmission*, this does not necessarily imply high-fidelity *copying* mechanisms (Saldana, Fagot, Kirby, Smith, & Claidière, 2019). We attribute this effect largely to collective cognitive alignment, meaning that cultural group members tend to perceive, remember, and reproduce information in consistent ways. To the extent that collective cognitive alignment is maintained through enculturation, whereby each individual "acquires" the cognitive biases of their group through interaction, it becomes possible that collective cognitive alignment may *fail* to be achieved either within or across generations. We have advanced cultural attractor theory by providing a socio-cognitive model of how cultural attractors may form, change, and stabilize in the absence of strongly-determining ecological constraints or innately-shared biases.

Our explorations with this model illustrate that factors at the scale of cognition, development, and demographic structure may interact in complex ways to shape patterns of collective cognitive alignment. First, we found that small amounts of noise in transmission may slow the rate of change of cultural attractor landscapes, promoting cognitive alignment within the population. In this way, noise at the level of perception and/or production may be counterintuitively beneficial for *reducing* errors at the level of cultural categories. Next, we found that longer learning times may result in a reduction of the number of categories at the population level over time, due to competition effects at the cognitive level. At the same time, critical periods of learning help to stabilize cultural attractor landscapes, because older agents remain in the population as "frozen" models for developing agents. Finally, we found that the complexity of cultural attractor landscapes decreases as population size grows larger, approaching a non-zero asymptote. This occurs because individuals in larger populations, in our baseline fully connected network, have fewer repeat interactions, which makes close alignment more difficult, and as a result, more diffuse categories are maintained in the population. This effect can be mitigated, however, through highly cliquish network structures that make repeat interactions very high, but this can come at the cost of global alignment. Our results offer a preview of the insights that may be gained by introducing more detailed representations of the culture-cognition feedback loop into more models of cultural evolution.

A crucial next step will be to include fitness constraints and selectionist transmission in our model. At present, the cultural attractors in our model are arbitrary and fitness-neutral. This

was an important simplification, since, as we have argued, the organic emergence of a cultural attractor landscape may be a critical *precondition* for Darwinian cultural selection, so an explanation of the emergence of cultural attractors must not ultimately fall back to Darwinian selection. Nonetheless, the cognitive capacities and developmental trajectories that facilitate the emergence of cultural attractors *are* themselves biologically evolved, so natural selection will need to be brought back into the picture in future work. Our model can be 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. For example, the cultural attractors that emerge could be mapped onto behaviors with immediate survival consequences, or onto frequency-dependent consequences such as when establishing shared systems of reference. We can also allow agents to generate *sequences* of signals, which may provide new insights into the entanglement between perceptual and combinatorial cognition in cultural attractor dynamics.

Integrating theories of Darwinian cultural selection with theories of cultural attraction—and theories of cognition more generally—will benefit from more mechanistic models of the feedback loop between cognitive development and population dynamics. Our model contributes to this theoretical bridge by representing cognitive, dyadic, developmental, and demographic dynamics simultaneously, in order to examine the conditions that either promote or inhibit the self-organization and maintenance of a stable cultural attractor landscape. Viewing cultural attractor landscapes as a complex system of interacting constraints at multiple levels allows for straightforward integration of cultural attractor theory with Darwinian selectionist accounts: fitness-based selection effects can be understood as yet another constraint on the formation of statistical attractor points. We hope this model will be useful for researchers interested in the co-evolution of innate cognitive biases, developmental tendencies, and demographic structure with culture.

Note

1 https://osf.io/6bsyx/?view_only=e91d9839ebe441a4841e3d312204e655

References

Acerbi, A., Charbonneau, M., Miton, H., & Scott-Phillips, T. (2019). Cultural stability without copying. Preprint. https://osf.io/vjcq3

Acerbi, A., Charbonneau, M., Miton, H., & Scott-Phillips, T. (2021). Culture without copying or selection. *Evolutionary Human Sciences*, *3*, E50.

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.

Acerbi, A., Mesoudi, A., & Smolla, M. (2020). *Individual-based models of cultural evolution. A step-by-step guide using R*. Routledge.

Ashby, F. G., & Maddox, W. T. (2005). Human category learning. *Annual Reviews in Psychology*, *56*, 149–178.

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, USA*, *107*(6), 2403–2407.

Bartlett, F. C., & Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge University Press.

Beckner, C., Blythe, R., Bybee, J., Christiansen, M. H., Croft, W., Ellis, N. C., Holland, J., Ke, J., Larsen-Freeman, D., et al,. (2009). Language is a complex adaptive system: Position paper. *Language Learning*, *59*, 1–26.

Bentz, C., & Winter, B. (2014). Languages with more second language learners tend to lose nominal case. In *Quantifying language dynamics* (pp. 96–124). Brill.

Birdsong, D. (1999). *Second language acquisition and the critical period hypothesis*. Routledge.

Boyd, R., & Richerson, P. J. (1988). *Culture and the evolutionary process*. University of Chicago Press.

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

Cantor, M., Chimento, M., Smeele, S. Q., He, P., Papageorgiou, D., Aplin, L. M., & Farine, D. R. (2021). Social network architecture and the tempo of cumulative cultural evolution. *Proceedings of the Royal Society B Biological Sciences*, *288*(1946), 20203107.

Cavalli-Sforza, L. L., & Feldman, M. W. (1973). Cultural versus biological inheritance: Phenotypic transmission from parents to children. (A theory of the effect of parental phenotypes on children's phenotypes). *American Journal of Human Genetics*, *25*(6), 618.

Cavalli-Sforza, L. L., & Feldman, M. W. (1981). *Cultural transmission and evolution: A quantitative approach*. Princeton University Press.

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.

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

Craver, C. F. (2006). When mechanistic models explain. *Synthese*, *153*(3), 355–376.

Creanza, N., Ruhlen, M., Pemberton, T. J., Rosenberg, N. A., Feldman, M. W., & Ramachandran, S. (2015). A comparison of worldwide phonemic and genetic variation in human populations. *Proceedings of the National Academy of Sciences*, *112*(5), 1265–1272.

Dale, R., & Lupyan, G. (2012). Understanding the origins of morphological diversity: The linguistic niche hypothesis. *Advances in Complex Systems*, *15*(3–4), 1150017.

Dawkins, R. (1976). *The selfish gene*. Oxford University Press.

Derex, M., & Boyd, R. (2016). Partial connectivity increases cultural accumulation within groups. *Proceedings of the National Academy of Sciences*, *113*(11), 2982–2987.

Enfield, N. J. (2014). *Natural causes of language: Frames, biases, and cultural transmission*. Language Science Press.

Epstein, J. M. (1999). Agent-based computational models and generative social science. *Complexity*, *4*(5), 41–60.

Fenk-Oczlon, G., & Pilz, J. (2021). Linguistic complexity: Relationships between phoneme inventory size, syllable complexity, word and clause length, and population size. *Frontiers in Communication*, *6*, 66.

Fields, C., & Levin, M. (2020). How do living systems create meaning? *Philosophies*, *5*(4), 36.

Fjell, A. M., & Walhovd, K. B. (2010). Structural brain changes in aging: courses, causes and cognitive consequences. *Reviews in Neurosciences*, *21*(3), 187–221.

Frankenhuis, W. E., & Panchanathan, K. (2011). Individual differences in developmental plasticity may result from stochastic sampling. *Perspectives on Psychological Science*, *6*(4), 336–347.

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

Gammaitoni, L., Neri, I., & Vocca, H. (2010). The benefits of noise and nonlinearity: Extracting energy from random vibrations. *Chemical Physics*, *375*(2-3), 435–438.

Gerkey, D. (2013). Cooperation in context: Public goods games and post-Soviet collectives in Kamchatka, Russia. *Current Anthropology*, *54*(2), 144–176.

Goldman, M. S. (2004). Enhancement of information transmission efficiency by synaptic failures. *Neural Computation*, *16*(6), 1137–1162.

Gong, T., Baronchelli, A., Puglisi, A., & Loreto, V. (2011). Exploring the roles of complex networks in linguistic categorization. *Artificial Life*, *18*(1), 107–121.

Hakuta, K., Bialystok, E., & Wiley, E. (2003). Critical evidence: A test of the critical-period hypothesis for second-language acquisition. *Psychological Science*, *14*(1), 31–38.

Healy, K. (2017). Fuck nuance. *Sociological Theory*, *35*(2), 118–127.

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. (2002). On modeling cognition and culture. *Journal of Cognition and Culture*, *2*(2), 87–112.

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

Heyes, C. (2018). *Cognitive gadgets: The cultural evolution of thinking*. Harvard University Press.

Hoehl, S., Keupp, S., Schleihauf, H., McGuigan, N., Buttelmann, D., & Whiten, A. (2019). 'over-imitation': A review and appraisal of a decade of research. *Developmental Review*, *51*, 90–108.

Horner, V., & Whiten, A. (2005). Causal knowledge and imitation/emulation switching in chimpanzees (*Pan troglodytes*) and children (*Homo sapiens*). *Animal Cognition*, *8*(3), 164–181.

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

Kalish, M. L., Griffiths, T. L., & Lewandowsky, S. (2007). Iterated learning: Intergenerational knowledge transmission reveals inductive biases. *Psychonomic Bulletin & Review*, *14*(2), 288–294.

Karmiloff-Smith, B. A. (1994). Beyond modularity: A developmental perspective on cognitive science. *European Journal of Disorders of Communication*, *29*(1), 95–105.

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.

Komarova, N. L., & Nowak, M. A. (2001). Natural selection of the critical period for language acquisition. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *268*(1472), 1189–1196.

Kuhl, P. K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Sciences*, *97*(22), 11850–11857.

Lazer, D., & Friedman, A. (2007). The network structure of exploration and exploitation. *Administrative Science Quarterly*, *52*(4), 667–694.

Lewis, H. M., & Laland, K. N. (2012). Transmission fidelity is the key to the build-up of cumulative culture. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, *367*(1599), 2171–2180.

Lewis, M. L., & Frank, M. C. (2016). The length of words reflects their conceptual complexity. *Cognition*, *153*, 182–195.

Lewontin, R. C. (1972). The apportionment of human diversity. In *Evolutionary Biology* (pp. 381–398). Springer.

Lupyan, G., & Dale, R. (2010). Language structure is partly determined by social structure. *PloS One*, *5*(1), e8559.

McClelland, J. L., Botvinick, M. M., Noelle, D. C., Plaut, D. C., Rogers, T. T., Seidenberg, M. S., & Smith, L. B. (2010). Letting structure emerge: connectionist and dynamical systems approaches to cognition. *Trends in Cognitive Sciences*, *14*(8), 348–356.

McClelland, J. L., Thomas, A. G., McCandliss, B. D., & Fiez, J. A. (1999). Understanding failures of learning: Hebbian learning, competition for representational space, and some preliminary experimental data. *Progress in Brain Research*, *121*, 75–80.

McDonnell, M. D., & Ward, L. M. (2011). The benefits of noise in neural systems: Bridging theory and experiment. *Nature Reviews Neuroscience*, *12*(7), 415–425.

Mesoudi, A. (2021). *Cultural evolution*. University of Chicago Press.

Mesoudi, A., & Whiten, A. (2004). The hierarchical transformation of event knowledge in human cultural transmission. *Journal of Cognition and Culture*, *4*(1), 1–24.

Miton, H., & Charbonneau, M. (2018). Cumulative culture in the laboratory: Methodological and theoretical challenges. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1879), 20180677.

Moran, S., McCloy, D., & Wright, R. (2012). Revisiting population size vs. phoneme inventory size. *Language*, 877–893.

Morin, O. (2013). How portraits turned their eyes upon us: Visual preferences and demographic change in cultural evolution. *Evolution and Human Behavior*, *34*(3), 222–229.

Morin, O. (2016). *How traditions live and die*. Oxford University Press.

Nowak, M. A., & Krakauer, D. C. (1999). The evolution of language. *Proceedings of the National Academy of Sciences*, *96*(14), 8028–8033.

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.

Panchanathan, K., & Frankenhuis, W. E. (2016). The evolution of sensitive periods in a model of incremental development. *Proceedings of the Royal Society B: Biological Sciences*, *283*(1823), 20152439.

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), 42–63.

Ravignani, A., Delgado, T., & Kirby, S. (2016). Musical evolution in the lab exhibits rhythmic universals. *Nature Human Behaviour*, *1*(1), 1–7.

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.

Revill, K. P., & Spieler, D. H. (2012). The effect of lexical frequency on spoken word recognition in young and older listeners. *Psychology and Aging*, *27*(1), 80.

Rulke, D. L., & Galaskiewicz, J. (2000). Distribution of knowledge, group network structure, and group performance. *Management Science*, *46*(5), 612–625.

Saldana, C., Fagot, J., Kirby, S., Smith, K., & Claidière, N. (2019). High-fidelity copying is not necessarily the key to cumulative cultural evolution: A study in monkeys and children. *Proceedings of the Royal Society B*, *286*(1904), 20190729.

Schläpfer, M., Bettencourt, L. M., Grauwin, S., Raschke, M., Claxton, R., Smoreda, Z., West, G. B., & Ratti, C. (2014). The scaling of human interactions with city size. *Journal of the Royal Society Interface*, *11*(98), 20130789.

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.

Smaldino, P. E. (2014). The cultural evolution of emergent group-level traits. *Behavioral and Brain Sciences*, *37*(3), 243–95.

Smaldino, P. E. (2017). Models are stupid, and we need more of them. In R. R. Vallacher, S. J. Read, & A. Nowak (Eds.), *Computational social psychology* (pp. 311–331). Routledge.

Smaldino, P. E. (2019). Social identity and cooperation in cultural evolution. *Behavioural Processes*, *161*, 108–116.

Smart, P. R., Huynh, T. D., Braines, D., & Shadbolt, N. (2010). Dynamic networks and distributed problem-solving. Knowledge Systems for Coalition Operations (KSCO'10), Vancouver, British Columbia, Canada. 20–22 Sep 2010. https://eprints.soton.ac.uk/271508/

Smolla, M., & Akçay, E. (2019). Cultural selection shapes network structure. *Science Advances*, *5*(8), eaaw0609.

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

Steels, L., & Belpaeme, T. (2005). Coordinating perceptually grounded categories through language: A case study for colour. *Behavioral and Brain Sciences*, *28*(4), 469-489.

Thompson, B., & Griffiths, T. L. (2021). Human biases limit cumulative innovation. *Proceedings of the Royal Society B*, *288*(1946), 20202752.

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.

McMurray, B., Aslin, R. N., & Toscano, J. C. (2009). Statistical learning of phonetic categories: Insights from a computational approach. *Developmental science*, *12*(3), 369–378.

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.

Von Uexküll, J. (1934). A stroll through the worlds of animals and men: A picture book of invisible worlds. *Semiotica*, *89*(4), 319–391.

Watts, D. J. (1999). Networks, dynamics, and the small-world phenomenon. *American Journal of sociology*, *105*(2), 493–527.

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

Wimsatt, W. C. (1972). Complexity and organization. In *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, volume 1972 (pp. 67–86). D. Reidel Publishing.

Winter, B., & Wedel, A. (2016). The co-evolution of speech and the lexicon: The interaction of functional pressures, redundancy, and category variation. *Topics in Cognitive Science*, *8*(2), 503–513.

Winters, J., Kirby, S., & Smith, K. (2015). Languages adapt to their contextual niche. *Language and Cognition*, *7*(3), 415–449.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.