The Cultural Brain Hypothesis: How culture drives brain expansion, sociality, and life history

Michael Muthukrishna[1,2*], Michael Doebeli[3], Maciej Chudek[4] & Joseph Henrich[2,5]

[1]Department of Psychological and Behavioural Science, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, United Kingdom

[2]Department of Human Evolutionary Biology, Harvard University, 11 Divinity Avenue, Cambridge, MA 02138, United States of America

[3]Department of Zoology / Department of Mathematics, University of British Columbia, Vancouver, BC, Canada

[4]School of Human Evolution and Social Change, Arizona State University, 900 Cady Mall, Tempe, AZ 85287, United States of America

[5]Canadian Institute for Advanced Research. Toronto, Canada

*Corresponding Author:

Email: m.muthukrishna@lse.ac.uk (MM)

Abstract

In the last few million years, the hominin brain more than tripled in size. Comparisons across evolutionary lineages suggest that this expansion may be part of a broader trend toward larger, more complex brains in many taxa. Efforts to understand the evolutionary forces driving brain expansion have focused on climatic, ecological, and social factors. Here, building on existing research on learning, we analytically and computationally model the predictions of two closely related hypotheses: The Cultural Brain Hypothesis and the Cumulative Cultural Brain Hypothesis. The Cultural Brain Hypothesis posits that brains have been selected for their ability to store and manage information, acquired through asocial or social learning. The model of the Cultural Brain Hypothesis reveals relationships between brain size, group size, innovation, social learning, mating structures, and the length of the juvenile period that are supported by the existing empirical literature. From this model, we derive a set of predictions—the Cumulative Cultural Brain Hypothesis—for the conditions that favor an autocatalytic take-off characteristic of human evolution. This narrow evolutionary pathway, created by cumulative cultural evolution, may help explain the rapid expansion of human brains and other aspects of our species' life history and psychology.

Author Summary

Humans have extraordinarily large brains, which tripled in size in the last few million years. Other animals also experienced a significant, though smaller, increase in brain size. These increases are puzzling, because brain tissue is energetically expensive—a smaller brain is easier to maintain in terms of calories. Here we present a theory, captured in an analytic and

computational model, that explains these increases in brain size: The Cultural Brain Hypothesis. The theory relies on the idea that brains expand to store and manage more information. Brains expand in response to the availability of information and calories. Information availability is affected by learning strategies (e.g. learning from others or learning by yourself), group size, mating structure, and the length of the juvenile period, which co-evolve with brain size. The model captures this co-evolution under different conditions and describes the specific and narrow conditions that can lead to a take-off in brain size—a possible pathway that led to the extraordinary expansion in our own species. We call these conditions the Cumulative Cultural Brain Hypothesis. These theories are supported by our tests using existing empirical data.

Introduction

In the last few million years, the cranial capacity of the human lineage dramatically increased, more than tripling in size [1-3]. This rapid expansion may be part of a gradual and longer-term trend toward larger, more complex brains in many taxa [3-7]. These patterns of increasing brain size are puzzling since brain tissue is energetically expensive [8-13]. Efforts to understand the evolutionary forces driving brain expansion have focused on climatic, ecological, and social factors [1-3, 14, 15]. Here we provide an integrated model that attempts to explain both the broader patterns across taxa and the human outlier. To do this, we develop an analytic model and agent-based simulation based on the Cultural Brain Hypothesis (CBH): the idea that brains have been selected for their ability to store and manage information via some combination of individual (asocial) or social learning [16-21]. That is, we develop the idea that bigger brains have evolved for more learning and better learning. The information acquired through these various learning processes is locally adaptive, on average, and could be related to a wide range of

behavioural domains, which could vary from species to species. The forms of learning we model could plausibly apply to problems such as finding resources, avoiding predators, locating water, processing food, making tools, and learning skills, as well as to more social strategies related to deception, coercion, manipulation, coordination or cooperation. Our theoretical results suggest that the same underlying selective process that led to widespread social learning [22] may also explain the correlations observed across species in variables related to brain size, group size, social learning, innovation, and life history. Moreover, the parameters in the formal representation of our theory offer hypotheses for why brains have expanded more in some lineages than others [4, 23].

Building on the Cultural Brain Hypothesis, our theoretical model also makes a set of predictions that we call the Cumulative Cultural Brain Hypothesis (CCBH). These predictions are derived from the parameters within the CBH model that favor an autocatalytic take-off in brain size, adaptive knowledge, group size, learning, and life history characteristic of human evolution. The CCBH has precedents in other models describing the processes that led to human uniqueness [see 18-20, 24, 25-29]. Since the CCBH is not a separate model, but instead additional predictions derived from the CBH model, this approach both seats humans within the broad primate spectrum created by the selection pressures we specify, and also accounts for our peculiarities and unusual evolutionary trajectory. That is, the same mechanisms that lead to widespread social learning can also open up a novel evolutionary bridge to a highly cultural species under some specific and narrow conditions—those specified by the CCBH. When these conditions are met, social learning may cause a body of adaptive information to accumulate over generations. This accumulating body of information can lead to selection for brains better at social learning as well as storing and managing this adaptive knowledge. Larger brains, better at social learning, then further foster the accumulation of adaptive information. This creates an autocatalytic feedback loop that enlists

sociality (population side and interconnectedness), social learning, and life history to drive up both brain size and adaptive knowledge in a culture-gene co-evolutionary duet—the uniquely human pathway. The juvenile period expands to provide more time for social learning. As biological limits on brain size are reached [e.g. due to difficulties in birthing larger brains, even in modern populations, see 30], increases in the complexity and amount of adaptive knowledge can take place through other avenues, such as division of information (and ultimately, division of labor), mechanisms for increasing transmission fidelity, such as compulsory formal schooling, and further expansion of the "adolescent" period between fertility and reproduction, spent in additional education (i.e. delayed birth of first child) [16]. This process modifies human characteristics in a manner consistent with more effectively acquiring, storing, and managing cultural information.

The CBH and CCBH are related, and can be explored with the same model, but we keep them conceptually distinct for two reasons. First, the cumulative culture-gene co-evolutionary process produces cultural products, like sophisticated multi-part tools and food processing techniques, that no single individual could reinvent in their lifetime [despite having a big brain capable of potent individual learning; 19]. The evolution of a second inheritance system—culture—is a qualitative shift in the evolutionary process that demands analyses and data above and beyond that required for the CBH. Second, it's possible that either one of these hypotheses could hold without the other fitting the evidence—that is, it might be the CCBH explains the evolutionary trajectory of humans, but the CBH doesn't explain the observed patterns in social learning, brain size, group and life history in primates (or other taxa); or, vice-versa.

Our approach is distinct, but related to the Social Brain Hypothesis [SBH; 31], which argues that brains have primarily evolved for dealing with the complexities of social life in larger groups (e.g., keeping track of individuals, Machiavellian reasoning, and so on). Initial evidence

112 supporting the SBH was an empirical relationship shown between social group size in primates
113 and some measure of brain size [different measures of brain size are typically highly correlated;
114 32]. Though this relationship does not hold outside the primate order, broader versions of the SBH
115 that encompass other aspects of social cognition have been informally proposed with
116 corresponding evidence from comparative studies. For example, a relationship has been shown
117 between brain size and regular association in mammalian orders [6, 33], mating structure in birds
118 and mammals [33], and social structure and behavioral repertoire in whales and dolphins [34].
119 Efforts to formally explore these ideas isolate four distinct evolutionary mechanisms. First,
120 McNally and collaborators have explored the Machiavellian arms race between cooperation and
121 deception [35, 36]. Second, Dávid-Barrett and Dunbar [37] simulate a relationship between
122 coordination costs and group size showing that more complex coordination (and therefore higher
123 cognitive complexity) is required as group size increases. Third, Gavrilets [38] models
124 collaborative ability as a proxy for socio-cognitive competencies, exploring the effect of between-
125 group selection and ecological pressures and showing that between-group competition can select
126 for collaborative ability, which is then further reinforced by ecological pressures. The predictions
127 of this last model are reinforced by a recent data-driven metabolic model tracking energy trade-
128 offs under different types of competition [39]. Finally, exploring a distinct additional mechanism,
129 Gavrilets and Vose [40] simulate an evolutionary competition among males for females in which
130 males can evolve larger brains with learning abilities that permit them to acquire more effective
131 strategies.

132 In his seminal paper, Humphrey [41] highlighted the importance of social learning, along
133 with several other social factors. The theory presented here is therefore consistent with this and
134 other early research that emphasized the learning aspects of the social brain [41, 42; for a more

recent discussion, see Whiten & van Schaik, 2007, 43, 44]. However, while many verbal descriptions of the SBH are general enough to encompass most aspects of the CBH, formal instantiations of the SBH each focus on quite distinct evolutionary mechanisms: (1) deception and cooperation, (2) coordination between group members, (3) cooperation in between-competition and against ecological challenges, and (4) learning social strategies. To make progress, we argue that it's crucial to distinguish the various evolutionary mechanisms that have often been clumped under the "social brain" rubric, and then test for the action of these various mechanisms (which need not be mutually exclusive).

The CBH and CCHB are a deliberate shift in focus from "social" to "learning"; a shift with precedence in other theories, most informally expressed [for example, see 15, 20, 21, 23, 45, 46]. There are, however, some clear departures from most previous approaches. First, crucial to this shift from social to learning is that group size evolves endogenously, rather than as a product of externalities (such as avoidance of predators). Second, learning is assumed to be more general than the skills and cognition required for social living. Individuals could learn skills and knowledge for social coordination, cooperation, and competition, such as social strategies to improve mating, as in [40]. But equally, these skills and knowledge may be related to other fitness relevant domains, such as ecological information about finding food or making tools. Indeed, the generality of adaptive knowledge is critical to the CCBH and the human take-off. In our approach, the potential for a runaway process to explain the human outlier arises neither from a Machiavellian arms race [35, 36] nor from sexual selection [40], but instead from the rise of cumulative cultural evolution as a second system of inheritance. Ecological factors are considered in the CBH in terms of survival returns on adaptive knowledge (e.g. easier acquisition of more calories or easier avoidance of predators, where easier means requiring less knowledge).

To further develop the CBH and CCBH, our models explore the interaction and coevolution of (1) learned adaptive knowledge and (2) genetic influences on brain size (storage/organizational capacity), asocial learning, social learning, and an extended juvenile period with the potential for payoff-biased oblique social learning (learning from members of the previous generation apart from biological parents). We explicitly model population growth and carrying capacity alongside genes and culture in order to theorize potential relationships between group size and other parameters, like brain size and adaptive knowledge, and also to examine the effects of sociality on the co-evolutionary process through two different parameters. We assume carrying capacity (though not necessarily population) is increased by the possession of adaptive knowledge (e.g., more calories, higher quality foods, better predator avoidance). Our model incorporates ecological factors and phylogenetic constraints by considering different relationships between birth/death rates and both brain size and adaptive knowledge. This allows us to formalize (and in particular, simulate) these evolutionary processes for taxa facing diverse phylogenetic and ecological constraints.

Models

We begin by laying out the key assumptions underlying both the analytical and simulation models. The predictions of the analytical model are derived using adaptive dynamics. We present the key insights that we are able to derive without the complexities of simulation. The full analytical model can be found in the Supplementary Material. We then build on the analytic solutions to fully explore the mechanisms underlying these insights using an agent-based evolutionary simulation. This simulation also allows us to explicitly track group size and relax some of our assumptions, allowing oblique learning, learning biases, and life history to evolve.

We present the key insights and predictions of our model in three ways. First, we explain the conditions under which we expect relationships between our variables and how the size of these relationships is affected by our parameters. In doing so, we verbally describe the core logic underlying the theory. Second, we compare our predictions to existing data, plotting our simulation results side-by-side with this existing data. If our predictions were inconsistent with existing empirical correlations, this would pose a significant challenge to our theory. Finally, we derive the Cumulative Cultural Brain Hypothesis predictions, laying out the narrow evolutionary regime under which an autocatalytic interaction between cultural and genetic inheritance is most likely to generate a human-like take-off.

Assumptions

Three key assumptions underlie our theory:

1. Larger and more complex brains are more costly than less complex brains because they require more calories, are harder to birth, take longer to develop, and have organizational challenges. Therefore, *ceteris paribus*, increasing brain size/complexity decreases an organism's fitness. For simplicity, we assume that brain size, complexity, and organization (e.g., neuronal density) are captured by a single state variable, which we will refer to as "size".
2. A larger brain correlates with an increased capacity and/or complexity that allows for the storage and management of more adaptive knowledge. Adaptive knowledge could potentially relate to locating food, avoiding predators, securing mates, processing resources (detoxification, increased calorie release), hunting game, identifying medicinal plants, making tools, and so on.

3. More adaptive knowledge increases an organism's fitness either by increasing its number of offspring compared to conspecifics and/or by reducing its probability of dying before reproduction. Adaptive knowledge can be acquired asocially, through experience, trial and error, and causal reasoning, or socially, by learning from others.

The logic that follows from these key assumptions is first formalized using an analytic approach—an adaptive dynamics evolutionary model [47], available in the Supplementary Materials. This model captures the logic and several of the key predictions of the CBH. We then simulate the logic to capture the co-evolutionary dynamics needed to generate the CCBH.

Analytical Model

The analytical adaptive dynamics model we present in the Supplementary Materials allows us to understand the evolution of brain size, adaptive knowledge, and reliance on social learning as a function of transmission fidelity, asocial learning efficacy, and survival returns on adaptive knowledge without the complexities of co-evolutionary dynamics and explicit evolution of oblique learning and learning biases. We can derive a set of predictions from the insights gained from this model.

Predictions

The key predictions from the analytical model are that:

1. Increased reliance on social learning requires high transmission fidelity (relative to the ability to generate knowledge by oneself).
2. Extreme reliance on social learning also assumes access to a range of models with different amounts of adaptive knowledge [determined by sociality - population size and interconnectedness - and assuming an ability to select and learn from models with more adaptive knowledge; see 16, 48, 49].

3. A greater return on adaptive knowledge (affected by λ; e.g. richness of environment) increases brain size (and may therefore explain different encephalization slopes across tax). Assuming an exponential return on adaptive knowledge, the environment will have a larger effect on social learners.

However, there are several assumptions and implications underlying these basic insights, such as:

1. Social learners face a bootstrapping problem of where the initial knowledge comes from.
2. The birth rate and the indirect relationships that affect actual population size will also affect brain size (and adaptive knowledge).
3. Species that do enter an extreme of social learning (such as humans) are on a treadmill, requiring higher transmission fidelity and more adaptive knowledge to sustain their large brains. A loss in either transmission fidelity or access to adaptive knowledge would drive the species towards smaller brains.

Brain size and reliance on social over asocial learning will depend on factors that affect availability of adaptive knowledge, which are themselves affected by learning strategies and adaptive knowledge. In other words, there are a range of co-evolutionary dynamics that we have assumed or abstracted away in order to solve this model analytically, but which are crucial to capture and understand the full range of evolutionary dynamics. To understand the conditions under which social learning might emerge (and perhaps more interestingly, extreme reliance on social learning as in humans), we need to explore these co-evolutionary dynamics. We explore these full set of variables and explore these dynamics through an evolutionary simulation. An

evolutionary simulation also allows us to properly account for population size, population structure, more sophisticated learning strategies, and life history. This model will bolster and expand on our analytic model and reveal the conditions where adaptive knowledge and brain size will increase.

Simulation Model

To explore the culture-gene co-evolutionary dynamics, we constructed an agent-based evolutionary simulation that extends our analytic model. In our simulation, individuals are born, learn asocially or socially from their parent with some probability, potentially update by asocial learning or by socially learning from more successful members of their group during an extended juvenile period, migrate between demes, and die or survive based on their brain size and adaptive knowledge. Individuals who survive this process give birth to the next generation. We are mainly interested in the effects of natural selection and learning, so we use a haploid model and ignore non-selective forces such as sex, gene recombination, epistasis, and dominance. The lifecycle of the model, as well as all variables and parameters, are shown in Figure 1 below.

Individual variables

b_{ij}: Brain size [0 - ∞]
a_{ij}: Adaptive knowledge [0 - ∞]
s_{ij}: Social learning probability [0 - 1]
v_{ij}: Oblique learning probability [0 - 1]
l_{ij}: Learning bias [-∞ - ∞]

Simulation Constants

ρ: Malthusian growth rate [0.8]
μ: Genetic mutation rate [0.005]
σ_b: Brain deviation [0.05]
σ_a: Adaptive knowledge deviation [0.1]
σ_s: Social learning deviation [0.1]
σ_o: Oblique learning deviation [0.1]
σ_l: Learning bias deviation [0.5]

Stage 1: BIRTH

Assumptions

1. Higher population adaptive knowledge can support a larger population
2. More individual adaptive knowledge leads to more offspring

Variables and parameters

φ: Reproductive skew (mating structure)
k_j: Carrying capacity of deme j
A_j: Mean adaptive knowledge in deme j
N_{jt}: Population in deme j at time t
χ: Scaling coefficient for k_j

See Equations 1 - 4

Stage 2: LEARNING

Variables and parameters

τ: Transmission fidelity
ζ: Asocial learning efficacy
A_j: Mean adaptive knowledge in deme j
N_{jt}: Population in deme j at time t

See Equations 5 - 6

Stage 3: MIGRATION

Variables and parameters

m: migration rate

Stage 4: SELECTION

Assumptions

1. Larger brains are more costly
2. Adaptive knowledge can mitigate the cost of a larger brain

Variables and parameters

c_{max}: Maximum brain cost
κ: Maximum brain empty brain
λ: Death rate mitigation (richness of ecology)

See Equation 7

262

Figure 1. Lifecycle of simulation. On the left we define all individual evolving variables and constants. Parameters are defined within the relevant life stage.

This simulation was written in C++ by MM (code in Supplemental Materials). To reduce bugs, two computer science undergraduate research assistants independently reviewed the code and wrote a suite of unit tests using Google's C++ Testing Framework. The simulation begins with 50 demes, each with a population of 10 individuals. Throughout the simulation, the number of demes was fixed at 50. In early iterations of the model, we explored increasing the number of demes to 100 for some of the parameter space and found no significant impact on the results. Our starting population of 10 individuals is roughly equivalent to a real population of 40 individuals, assuming two sexes and one offspring per parent (4×10). As a reference, mean group size in modern primates ranges from 1 to 70 [32].

Each individual i in deme j has a brain of size b_{ij} with a fitness cost that increases with increasing brain size. Adaptive knowledge is represented by a_{ij}, where $0 \leq a_{ij} \leq b_{ij}$. Increasing adaptive knowledge can mitigate the selection cost of a larger brain, but such knowledge is limited by brain size.

Our simulations begin with individuals who have no adaptive knowledge, but the ability to fill their $b_{ij} = 1.0$ sized brains with adaptive knowledge through asocial and/or social learning with some probability. To explore the idea that juvenile periods can be extended to lengthen the time permitted for learning, we included two stages of learning. In both learning stages, the probability of using social learning rather than asocial learning is determined by an evolving *social learning probability* variable (s_{ij}). We began our simulations with the social learning probability variable set to zero (i.e. at the beginning of the simulation, all individuals are asocial learners). To

explore the invasion of asocial learners into a world of social learners, we also ran the simulation with the social learning probability variable set to one (i.e. at the beginning of the simulation, all individuals are social learners). Although social learning is widespread in the animal kingdom [22], a realistic starting point is closer to pure asocial learning. Nevertheless, the simulations starting with social learners were often useful in understanding these dynamics, so, in some cases where it is insightful, we report these results as well.

Asocial learning allows for the acquisition of adaptive knowledge, independent of the adaptive knowledge possessed by other individuals. In contrast, social learning allows for vertical acquisition of adaptive knowledge possessed by the genetic parent in the first learning stage or oblique acquisition from more knowledgeable members of the deme (from the parental generation) in the second learning stage. The tendency to learn from models other than the genetic parent is determined by a genetically evolving *oblique learning probability* variable (v_{ij}). Thus, the simulation does not assume oblique learning or a second stage of learning [a misplaced critique of related models in our opinion; 50; but a critique not relevant to the present model, 51]. The probability of engaging in a second round of oblique social learning is a proxy for the length of the juvenile period. In the second stage of learning, if an individual tries to use social learning, but does not use oblique learning, no learning takes place beyond the first stage. This creates an initial advantage for asocial learning and cost for evolution to extend learning into an extended juvenile period. We also allow the ability to select a model with more adaptive knowledge (for oblique learning) to evolve through a *payoff-bias ability* variable (l_{ij}).

These simulations result in a series of predicted relationships between brain size, group size, adaptive knowledge, asocial/social learning, mating structure, and the juvenile period. Some of these relationships have already been measured in the empirical literature and thus provide

immediate tests of our theory. Specifically, several authors have shown positive relationships (notably in primates) between (1) brain size and social group size [4, 31, 52], (2) brain size and social learning [46, 53], (3) brain size and length of juvenile period [54-57], and (4) group size and the length of the juvenile period [56].

Various hypotheses have been proposed for these relationships. Here we argue that they are all a consequence of a singular evolutionary process, the dynamics of which the CBH models reveal. In addition, we find that different rates of evolutionary change and the size of these relationships across taxa [6] may be accounted for by the extent to which adaptive knowledge reduces the death rate (λ in our model). This λ term captures any factor that moderates the relationship between adaptive knowledge and survival. One interpretation, but by no means the only one, is the resource richness of the ecology. For example, richer ecologies offer more 'bang for the buck', that is, more calories unlocked for less knowledge, allowing individuals to better offset the size of their brains. Higher λ suggest a richer ecology—or more specifically, an ecology where smarts have a greater return on survival. Indeed, research among primates has revealed that factors affecting access to a richer ecology—home range size or the diversity of food sources—are associated with brain size [58, 59]. Thus, our model may help explain why both social and ecological variables seem to be variously linked to brain size.

The dynamics of our model also reveal the ecological conditions, social organization and evolved psychology most likely to lead to the realm of cumulative cultural evolution, the pathway to modern humans. These predictions capture the CCBH. Our model indicates the following pathway. Under some conditions, brains will expand to improve asocial learning and thereby create more adaptive knowledge. This pool of adaptive knowledge leads to selection favoring an immense reliance on social learning, with selective oblique transmission, allowing individuals to

exploit this pool of growing knowledge. Rogers' [60] paradox, whereby social learners benefit from exploiting asocial learners' knowledge, but do not themselves generate adaptive knowledge, is solved by selective oblique social learning transmitting accidental innovations to the next generation. Under some conditions, an interaction between brain size, adaptive knowledge, and sociality (deme size and interconnectedness) emerges, creating an autocatalytic feedback loop that drives all three—the beginning of cumulative cultural evolution.

The Lifecycle. Individuals go through four distinct life stages (see Figure 1): Individuals (1) are born with genetic traits similar to their parents, with some mutation, (2A) learn adaptive knowledge socially from their parents or through asocial learning independent of their parents, (2B) go through a second stage of learning adaptive knowledge through asocial learning or oblique social learning, (3) migrate between demes, and (4) die or survive to reproduce the next generation. Fecundity and viability selection (birth and death) are expressed separately, allowing us to disentangle the effect of adaptive knowledge on outcompeting conspecifics and on reducing the risk of dying before reproduction.

Stage 1: The Birth Stage

In the birth stage, the individuals who survive the selection stage (Stage 4) give birth to the next generation.

Adaptive Knowledge and the Number of Offspring. We assume that demes with greater mean adaptive knowledge can sustain a larger population. We formalized this assumption in Equation 1 by linking k_j, which affects the carrying capacity of the deme, to the mean adaptive knowledge of the individuals in the deme (A_j) and some minimum value that we set to our starting group size ($N_{j_0} = 10$). The relationship between mean adaptive knowledge and k_j is scaled by χ, but adjusting this coefficient resulted in a computationally intractable deme size as adaptive

knowledge accumulated. Therefore, we set this coefficient to a constant value ($\chi = 10$) and left exploration of this parameter for a future model. The deme size (N_{j_t}) in the current generation (t) and k_j are then used to calculate the total expected number of offspring ($N_{j_{t+1}}$) in the next generation ($t + 1$) using the discrete logistic growth function in Equation 2, where ρ is the generational growth rate. Initial simulations suggested that ρ only affected the rate of evolution rather than the qualitative outcomes. We selected a reasonable value ($\rho = 0.8$) based on Pianka (61).

$$k_j = \chi A_j + N_{j_0} \quad (1)$$

$$N_{j_{t+1}} = \frac{N_{j_t} e^{\rho}}{1 + \left(\frac{N_{j_t}}{k_j}(e^{\rho} - 1)\right)} \quad (2)$$

Equation 2 tells us the Expected Value for the number of offspring based on current deme size and k_j (based on deme mean adaptive knowledge). However, this does not tell us which individuals within the deme gave birth to the offspring. We assume that more adaptive knowledge increases an individual's birth rate. We parameterized the strength of the relationship between adaptive knowledge and birth rate (fecundity selection). A potential parent's (i_j) probability of giving birth (p_{ij}) is given by their sigmoid transformed adaptive knowledge value (Equation 3) as a fraction of the sum of all transformed adaptive knowledge values of individuals in the deme (Equation 4). The transformation is adjusted by φ, allowing us to study the importance of fecundity selection. For example, we can turn off fecundity selection entirely by setting $\varphi = 0$: A world with no reproductive skew; all potential parents have the same probability of giving birth. The more we increase φ, the more we have a winner-takes-all world, where to win, one has to acquire adaptive knowledge. This is crucial in thinking about how, for example, our culture-gene co-evolutionary

process is influenced by social organization and mating structures that create high reproductive skew.

The φ parameter affected reproductive skew by increasing the breeding bias toward those with more adaptive knowledge. Though mating structure and reproductive skew are separable concepts, increased pair-bonding correlates with reduced reproductive skew. Thus mating structure is one mechanism, though not the only mechanism, that may affect reproductive skew. A perfectly monogamous pair-bonded society with no differential selection at the birthing stage would have $\varphi = 0$. Increasing φ allows for an increase in polygyny from "monogamish" (mostly pair-bonded) societies at low values of φ to highly polygynous winner-takes-all societies where males with the most adaptive knowledge have significantly more offspring (see Figure 2). Our model suggests that in these high reproductive skew societies, such as more polygynous societies, variation is reduced. This allows for the initial rapid evolution of larger brains, but with little or no variation, populations are unable to use social learning to increase their adaptive knowledge and are more likely to go extinct. At the other extreme, evolutionary forces are quashed when $\varphi = 0$. Social learning and the advent of culture-gene coevolution are more likely to occur when reproductive skew is supressed, such as in monogamish or cooperative/communal breeding societies or where sharing norms result in shared benefits despite skew in ability or success [see 62, 63, 64]. Of course, some argue that culture supports, or is responsible for, such mating structures in humans, which would require us to endogenize φ. In the present model, we treat φ as a parameter.

Migration was fixed at 10% and thus, φ also affected the relative strength of individual, within-group selection and between-group selection. Between-group selection dominates when $\varphi = 0$ and is reduced as $\varphi \gg 0$.

$$a_{ij}^{T} = \frac{1}{1 + e^{-\varphi(a_{ij} - A_j)}} \quad (3)$$

$$p_{ij} = \frac{a_{ij}^{T}}{\sum_{i=1}^{N_j} a_{ij}^{T}} \quad (4)$$

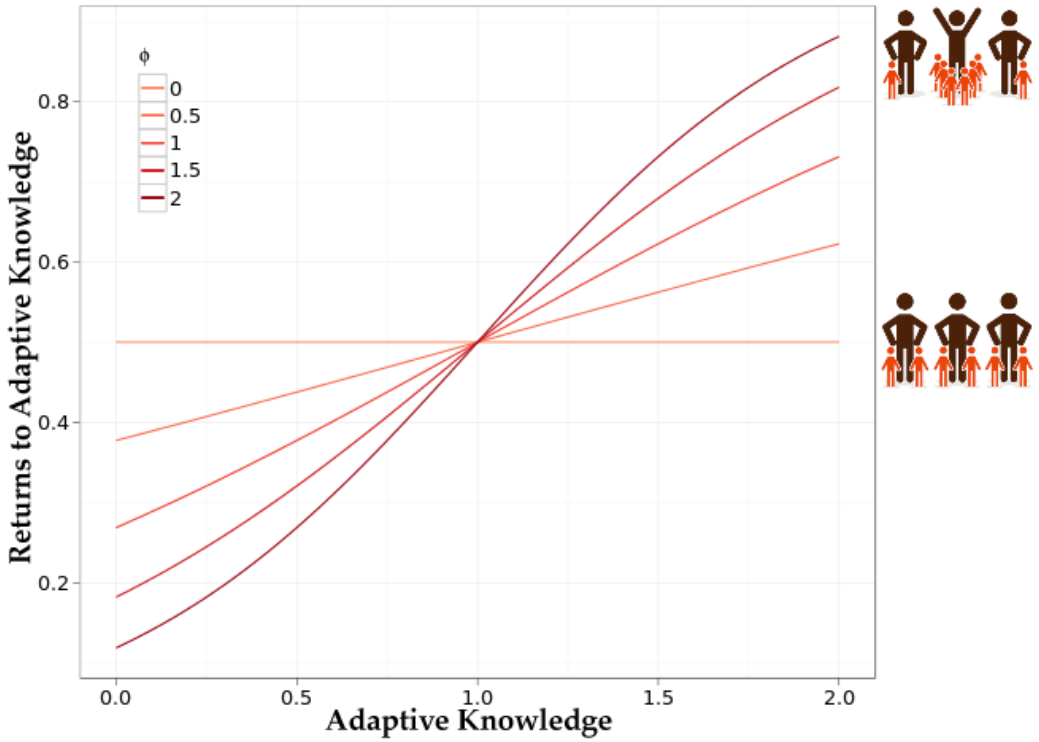


Figure 2. The effect of φ on transforming adaptive knowledge. Here the mean adaptive knowledge of the deme is 1 ($A_j = 1$).

We assume that more individual adaptive knowledge (a_{ij}^{T}) is associated with increased relative fertility. Using a binomial distribution, we instantiate the expected number of offspring n_{ij} for each parent. A binomial distribution $\mathrm{B}(n, p)$ describes the number of successes in a sequence of n binary experiments (in our model, have offspring vs. don't have offspring). The probability of success in any particular 'coin flip' is given by p. For each parent, we draw a value from a binomial distribution where the number of experiments is the Expected Value for the

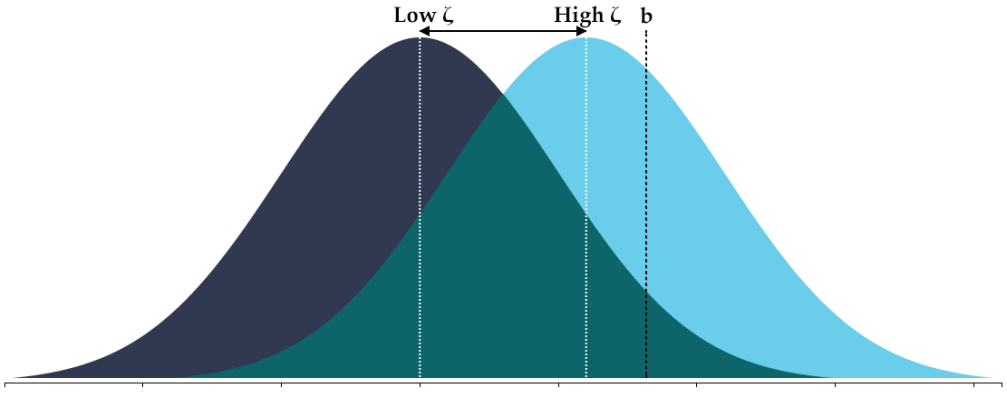
number of offspring in the deme ($n = N_{j_{t+1}}$) and the probability is calculated by Equation 16, i.e. from $\mathrm{B}\left(N_{j_{t+1}}, p_{ij}\right)$. By drawing these values from a binomial distribution, the sum of Expected Values for the offspring of all parents is $N_{j_{t+1}}$ (i.e. $N_{j_{t+1}} = \sum_{i=1}^{N_j} E\left[\mathrm{B}\left(N_{j_{t+1}}, p_{ij}\right)\right] = E\left[\sum_{i=1}^{N_j} n_{ij}\right]$).

Genetic Transmission and Mutation. The offspring (designated by a prime symbol) born to a parent are endowed with genetic characteristics similar to their parents. These offspring acquire four genetic traits from their parents—their brain size (b'_{ij}), social learning probability (s'_{ij}), oblique learning probability (v'_{ij}), and oblique learning bias (l'_{ij}). For each trait, newborn individuals have a $1-\mu$ probability of having the same value as their parents ($b_{ij}, s_{ij}, v_{ij}, l_{ij}$). If a mutation takes place, new values are drawn from a normal distribution with a mean of their parent value and a standard deviation σ_s for s'_{ij}, σ_v for v'_{ij}, σ_l for l, and σ_v for v'_{ij} and $\sigma_b b_{ij}$ for b'_{ij}. The standard deviations of s'_{ij} and v'_{ij} are not scaled by the mean, since these are probabilities and therefore bounded [0,1] (the normal distribution is truncated at [0,1]). Although l'_{ij} is not bounded, we do not scale the standard deviation by the mean, because small changes in l'_{ij} have a large effect on learning bias, due to the sigmoid function. Once offspring have been endowed with genetic characteristics, they then acquire adaptive knowledge. Their method and ability to acquire adaptive knowledge is affected by their genetic traits.

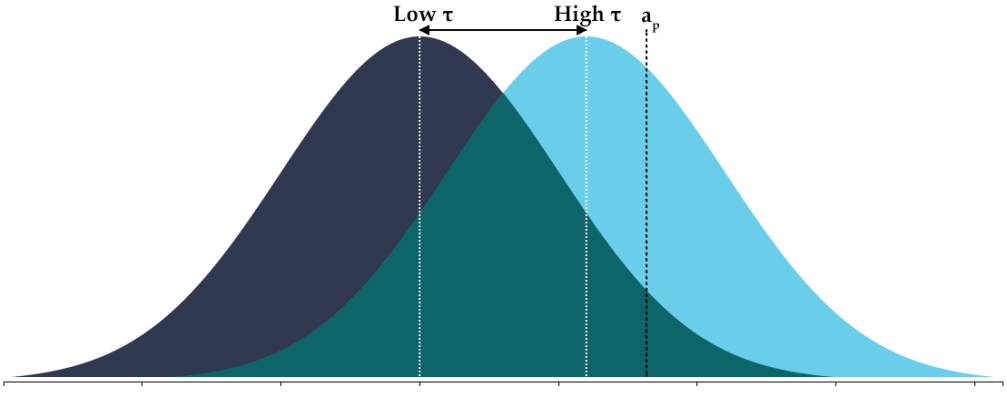
Stage 2: Learning

Asocially learned adaptive knowledge values (a'_{ij}) are drawn from a normal distribution based on an individual's brain size: $N\left(\zeta b'_{ij}, \sigma_a \zeta b'_{ij}\right)$. Rather than fix the variance and imply that the space of deviation in learning remains the same regardless of what has been learned, we allow the variance to scale with the mean of the distribution reflecting the idea of a *thought space* [16],

where the space of possible deviations grows as the amount of knowledge grows. Socially learned adaptive knowledge values are drawn from a similar normal distribution, but with a mean of the model's (t) adaptive knowledge value scaled by transmission fidelity (τ): $N\left(\tau a_{tj}, \sigma_a \tau a_{tj}\right)$ and the variance similarly scaled by the mean. Figure 3 below illustrates the distributions from which these values are drawn and the effect of ζ and τ.



(a)



(b)

Figure 3. Illustration of distributions for how asocial learning and social learning acquire adaptive knowledge. In (a) an asocial learner has a higher probability of drawing a value closer to their brain size if ζ is higher. In (b) a social learner has a higher probability of

drawing a value closer to their model's adaptive knowledge value if τ is high. Note that in both cases, adaptive knowledge cannot exceed brain size ($a_{ij} \leq b_{ij}$).[i]

For both asocial and social learning, an individual's adaptive knowledge may not exceed their brain size. But, compared to social learning, asocial learning enables the immediate acquisition of adaptive knowledge based on one's own brain size. Social learning is dependent on the adaptive knowledge possessed by parents, or those in the parents' generation within the same deme, if selection extends the learning phrase through a juvenile period.

In Stage 2A, newborn individuals i'_j can socially acquire adaptive knowledge from their parent i with probability s'_{ij}. If newborns do not learn from their parents $(1 - s'_{ij})$, they learn asocially instead.

In Stage 2B, individuals i'_j may update their adaptive knowledge through asocial learning with probability $(1 - s'_{ij})$ in the same manner as Stage 2A or obliquely from non-parents with probability $s'_{ij}v'_{ij}$. Individuals who do not asocially learn nor obliquely learn do no further learning. This allows us to study conditions under which oblique learning emerges during this extended learning period. Crucially, oblique learning has to out-compete a second round of asocial learning.

We adjust the strength of the relationship between a potential model's (m) adaptive knowledge and their likelihood of being modeled using the learner's l'_{ij} variable in the sigmoid tranformation function (5). A potential model's (t_j) probability of being selected (p_{tj}) is given by (6). Notice that these have the same functional form as Equations 3 and 4, and thus the transformation is similar to Figure 2. Both asocial and social learning only update adaptive knowledge values if these values are larger than those acquired during the first stage of learning, Stage 2A.

$$a_{mj}^{T} = \frac{1}{1 - e^{-l_{ij}(a_{mj} - A_j)}} \quad (5)$$

$$p_{mj} = \frac{a_{mj}^{T}}{\sum_{i=1}^{N_j} a_{mj}^{T}} \quad (6)$$

Note, since we are interested in the evolution of social learning, we stacked the deck somewhat against social learning. Individuals have a $s'_{ij} - s'_{ij}v'_{ij}$ chance of not doing *any* learning during Stage 2B. This creates an initial disadvantage for social learning, since any selection for social learning in Stage 2A risks missing out on a second round of asocial learning in Stage 2B.

Stage 3: Migration

Individuals migrate to a randomly chosen deme (not including their own) with probability $m = 0.1$, fixed to reduce the number of parameters. All demes have the same probability of immigration. Individuals retain their adaptive knowledge and genetic traits. There is no selection during migration; all individuals survive the journey.

Stage 4: Selection Based on Brain Size and Adaptive Knowledge

We formalized the assumption that larger, more complex brains are also more costly using a quadratic function to link brain size to maximum death rate (c_{max}), capturing the idea that the costs of large brains escalate non-linearly with size. In early simulations, we also tested an exponential function, but our exploration revealed no important qualitative differences between the functions.

To formalize the assumption that individuals with more adaptive knowledge are less likely to die *ceteris paribus*, we use the negative exponential function in Equation 7. The λ parameter in Equation 7 was varied between simulations and was used to determine the extent to which adaptive knowledge can offset the costs of brain size, where $\lambda = 0$ indicates no offset. As in our analytical

model, the λ parameter can be interpreted as how much adaptive knowledge one requires to unleash fitness-enhancing advantages.

$$d'_{ij} = c_{max} e^{-\lambda \frac{a'_{ij}}{b'_{ij}}} \quad (7)$$

This function captures the idea that the increasing costs of big brains can be offset by more adaptive knowledge. We set $c_{max} = \beta b^2$; $\beta = 1/10000$ in our simulation). This results in a maximum empty brain size of $b = 100$. The choice of setting the maximum empty brain size to $b = 100$ was somewhat arbitrary, but allowed for a reasonable size brains to see a range of evolutionary behavior (it just sets the scaling). We illustrate the effect of λ in Figure 4 below.

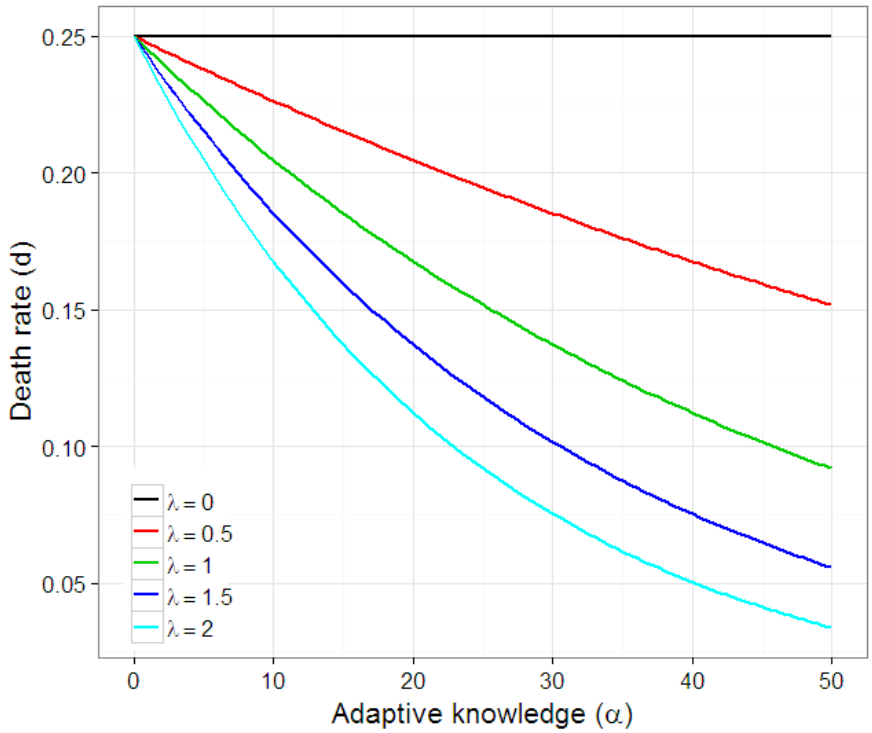


Figure 4. Reduction in death rate for different values of λ for a given brain size ($b = 50$ in this example).

Summary

These basic assumptions generate conflicting selection pressures for (1) more adaptive knowledge and (2) smaller brains. Under some conditions, the cost of having a larger brain is offset by the increased knowledge capacity of larger brains. If adaptive knowledge were freely available, there would be no constraint on the co-evolution of brains and adaptive knowledge; both would ratchet upward. In general, three related constraints prevent this from happening:

1. Adaptive knowledge does not always exist in the environment to fill a larger brain.
2. Larger brains without adaptive knowledge are costly without any offsetting benefits. This is especially true for social learners with brains larger than their parents, since this additional brain space cannot immediately be utilized.
3. Increases in brain size show diminishing returns; brain costs increase at a greater than linear rate.

We simulated a range of space within each parameter set for low, middle, and high values of other parameters for which we found interactions and realistic values of all other parameters. The range for each parameter was as follows: $\varphi[0.0,1.0]$, $\tau[0.75,1.0]$, $\zeta[0.1,0.9]$, $m[0.0,0.2]$, and $\lambda[0.0,2.0]$.

To give our populations enough time to evolve, we ran our simulation for 200,000 generations. Assuming 25-30 years per generation [66], this represents 5-6 million years of evolution, approximately the time since the hominin split from chimpanzees [67]. With a few exceptions, this guarantees that our genetically evolved variables have hit quasi-equilibrium. To account for stochastic variation in simulation outcomes, we performed 5 iterations per set of unique parameters and averaged the results across these. Unlike the other parameters, learning bias l did not generally reach equilibrium; however, we would not expect it to do so since higher l values

continue to provide an advantage in selecting models, such that l should slow down but continue to approach ∞. In our model, l is a one-dimensional state variable that captures better and worse ability to select models, but of course in the real world, there are a range of strategies and biases that have evolved to solve the problem of selecting models with more adaptive knowledge. For a discussion of the evolution of these biases and strategies and the trade-offs between them, see [19, 68]. For a list of such biases and strategies, see [69, 70].

Results

We begin by discussing the underlying processes that have led to the relationships between brain size, group size, social learning, and life history observed in the literature. We discuss the effect of our different parameters in creating these relationships and driving evolutionary patterns.

To benchmark the predictions derived by our model, we treat the quasi-equilibrium outcomes of each of our simulation runs as "quasi-species", with state variables representing the characteristics of each species. We qualitatively compare these simulation outcomes to existing empirical findings in the literature. Then, we focus on the CCBH and examine the conditions that favor substantial amounts of cumulative cultural evolution. The goal here is to understand the conditions under which the interaction between social learning, brain size, group size, sociality and life history generates the kind of auto-catalytic take-off required to explain the last two million years of human evolution.

The Cultural Brain Hypothesis

Overall, our evolutionary simulations produce patterns that are consistent with the existing empirical data, though, of course, our simulation produces many patterns that have not yet been examined. The causal relationships underlying these patterns—the CBH and our simulated

535 instantiation of it—are outlined in Figure 5 below. Before digging into the details, we summarize
536 these relationships as follows:

537 1. Larger brains allow for more adaptive knowledge. More adaptive knowledge can, in turn,
538 exert a selection pressure for larger brains.
539 2. More adaptive knowledge allows for larger potential carrying capacity. Consistent with our
540 analytical model, when there is sufficient adaptive knowledge and transmission fidelity is
541 high enough, there is selection for social learning to take advantage of the adaptive
542 knowledge; larger groups produce more adaptive knowledge that can be exploited by those
543 with better social learning abilities.
544 3. Large groups of individuals who primarily rely on social learning have larger bodies of
545 knowledge than those who rely on asocial learning, exerting a selection pressure for an
546 extended juvenile period in which more adaptive knowledge can be learned (and created).
547 4. An extended juvenile period (e.g. adolescence) is a period of reliance on oblique learning
548 (learning from non-genetic parents in the group), which creates a selection pressure for
549 learning biases better able to select individuals and knowledge to learn (better learning
550 abilities and tendency to learn from non-genetic models reinforce each other in a world of
551 plentiful and accumulating adaptive knowledge).
552 5. Oblique learning and learning biases lead to the realm of cumulative cultural evolution.
553 The length of the juvenile period (period between weaning and sexual maturity) varies
554 across species [56, 57], but adolescence (period between sexual maturity and reproduction)
555 may be uniquely human (possible exceptions include elephants [71] and orca [72]).
556 Adolescence may represent a period of oblique social learning, a key to cumulative cultural
557 evolution.

The Cultural Brain Hypothesis predicts that brain size, group size, adaptive knowledge, and the length of juvenile period should be positively intercorrelated among taxa with greater dependence on social learning, but are generally weaker or non-existent among taxa with little social learning. There has been less empirical data published for species with little social learning, perhaps due to a bias toward only publishing statistically significant relationships. The lack of this data makes it difficult to test the asocial regime predictions.

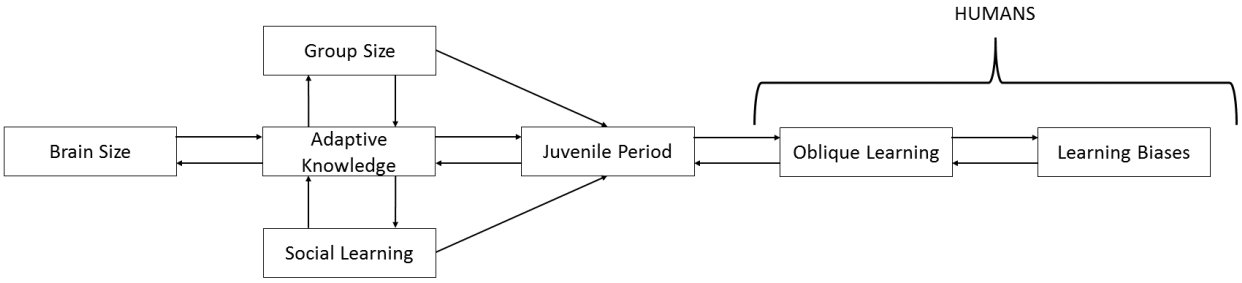


Figure 5. Here we illustrate the causal relationships predicted by the Cultural Brain Hypothesis. Larger brains allow for the storage and management of more information. More adaptive knowledge supports larger brains and larger groups. Larger groups possess more adaptive knowledge for social learning to exploit. Sufficiently large groups of social learners with sufficient knowledge create a selection pressure for a longer juvenile period for social learners to acquire knowledge selectively via biased oblique learning.

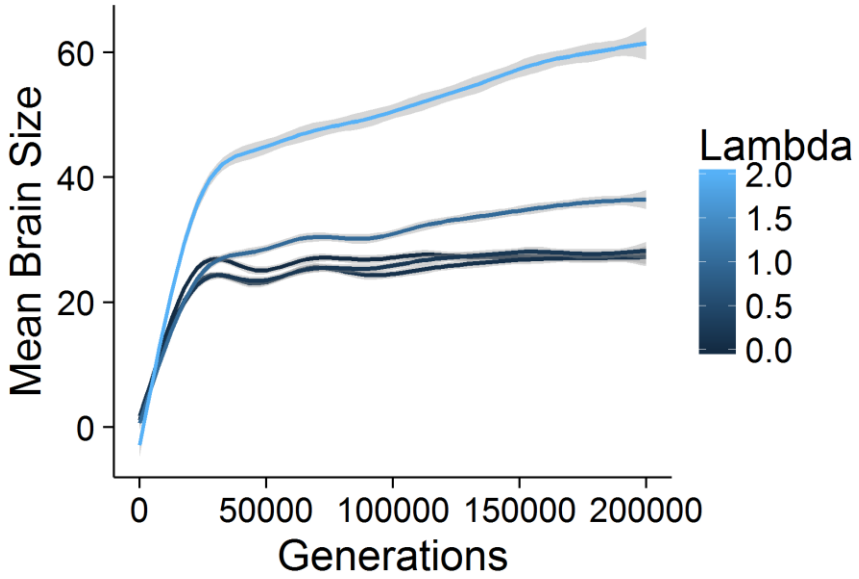
The strength of these relationships, overall brain size, and the evolution of different regimes vary, depending on the other parameters in our model. These include ecological factors such as the the strength of the relationship between adaptive knowledge and survival (λ), which we will call the "richness of the ecology" as a shorthand, as well as other factors that are themselves products of evolution (which we've held fixed as phylogenetic constraints): the relationship between adaptive knowledge and relative reproductive payoffs (φ), which are related to reproductive skew,

mating structure, and the level of individual vs between-group selection (we will refer to this as reproductive skew as a shorthand); transmission fidelity (τ); and asocial learning efficacy (ζ). Other models have theorized the evolution of these structures, tendencies, and abilities, but here we are interested in the effect of these factors on the co-evolutionary processes shown in Figure 5.

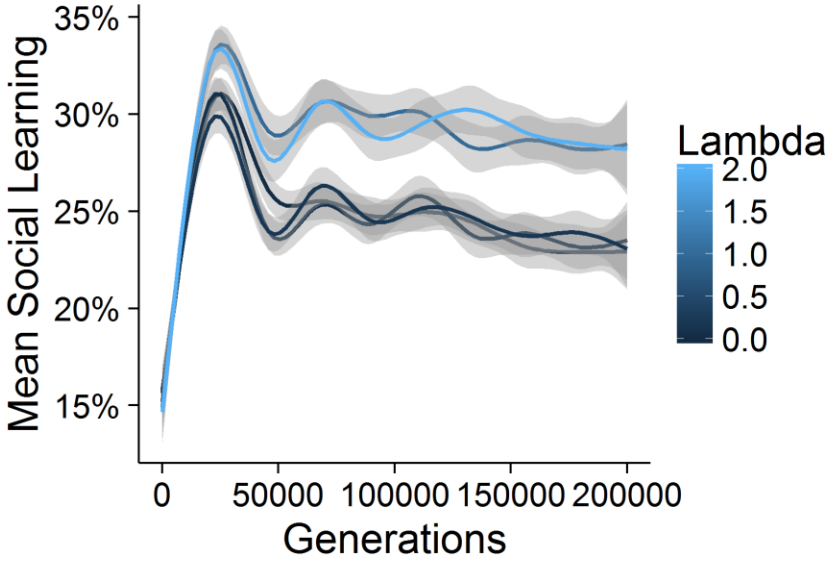
Effect of Parameters

Richness of the Ecology ($\boldsymbol{\lambda}$)

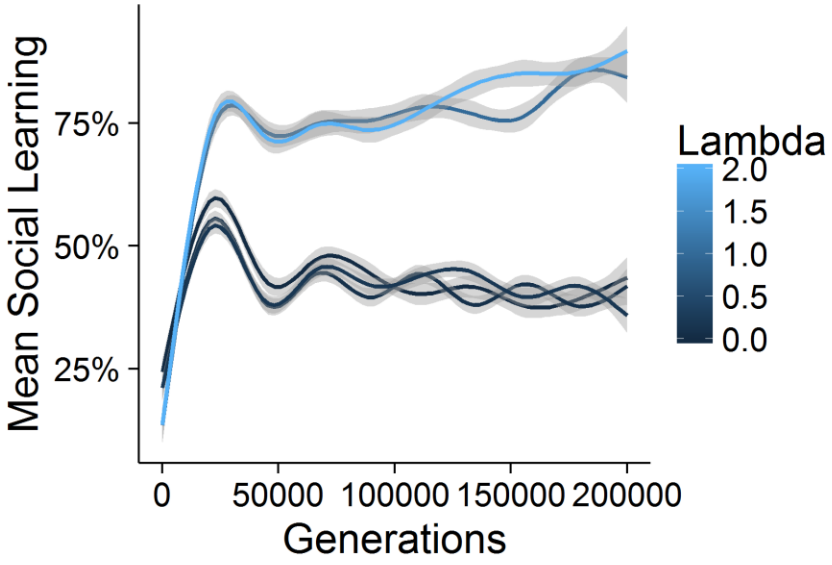
Our simulation suggests that the richness of the ecology (or at least greater returns for less knowledge) may be one factor that predicts both the rate of brain evolution and sociality. In a rich ecology (higher λ), less adaptive knowledge is needed to unlock more calories, navigate environmental hazards, evade more predators, and so on, allowing for larger brains; i.e. when λ is higher, adaptive knowledge offers more "bang for the buck". For those in the realm of social learning, in richer ecologies, we see greater reliance on social learning and larger brains (see Figure 6). Thus the CBH suggests that the empirical correlation that has been shown between sociality and the differential rate of brain expansion between taxa [6] may be explained by a third variable: the relationship between adaptive knowledge and survival.


Mean Brain Size
60
40
20
0
0
50000
100000
150000
200000
Generations
Lambda
2.0
1.5
1.0
0.5
0.0

(a)


Mean Social Learning
35%
30%
25%
20%
15%
0
50000
100000
150000
200000
Generations
Lambda
2.0
1.5
1.0
0.5
0.0

(b)



(c)

Figure 6. Here we show the effect of richness of the ecology on brain size and social learning. These are aggregated over a range of other parameters (a) Mean brain size showing the encephalization slope for different values of λ. Richer ecologies have a steeper slope for brain evolution. (b) Mean social learning showing the slopes over time. Richer ecologies support more social learning when social learning is adaptive. (c) This is made clear in the same plot for a narrower range of other parameters ($\tau = 1$ and $\zeta = 0.7$).

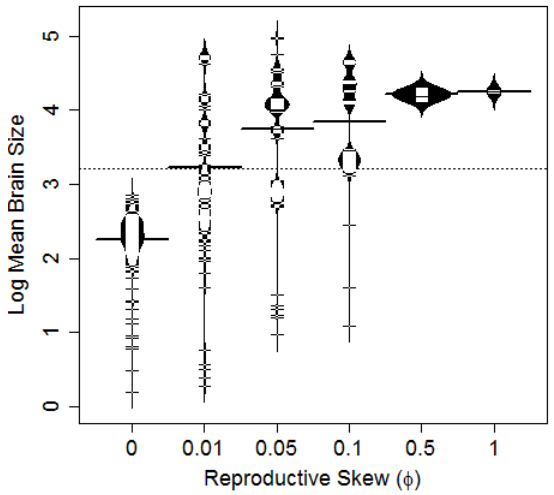
Reproductive Skew ($\boldsymbol{\varphi}$)

We model the effect of reproductive skew (which is related to mating structure and which also affects the relative individual-level to between-group-level selection) using φ. The φ parameter affects the relationship between individual adaptive knowledge and the mating competition. When $\varphi = 0$, all individuals have the same probability of reproducing regardless of their adaptive knowledge. There are different sources of reproductive skew. As an example, we use polygyny to illustrate the effect of this parameter. $\varphi = 0$ corresponds to a perfectly

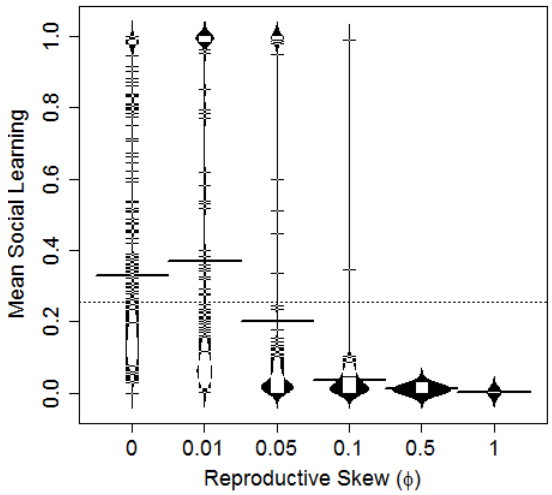
612 monogamous society with no fecundity selection, where all individuals have roughly the same
613 number of children. As φ increases, we enter into a slightly 'monogamish' or human cooperative
614 breeding society [or any other conditions where reproductive skew is limited; 62] and then to a
615 polygynous society (or at least a society where very few individuals have many children, and many
616 individuals have very few children) for very high values of φ. Increasing φ, increases the strength
617 of individual selection for more adaptive knowledge, but the results of this increase in fecundity
618 selection may be surprising.

619 First, brain size increases with φ (Figure 7a), but this relationship is misleading, because
620 the extinction rate (defined as the percentage of all individuals who die and do not pass their genes
621 or adaptive knowledge) also increases with higher φ (Figure 7c). Extinction rates go up, because
622 variance in brain size and the variance in the tendency to use social learning is reduced with too
623 high fecundity selection, as can be seen in Figures 7a and 7b. Here, more adaptive knowledge is
624 sought at any cost, but in a world with little adaptive knowledge, the best way to acquire this
625 knowledge is via asocial learning. This leads to populations getting stuck in the world of
626 homogenous larger brained asocial learners without the necessary variance in social learning
627 ability or proclivity (attendance and learning from conspecifics) to take advantage of the existing
628 body of adaptive knowledge and transition to smaller brained social learners.

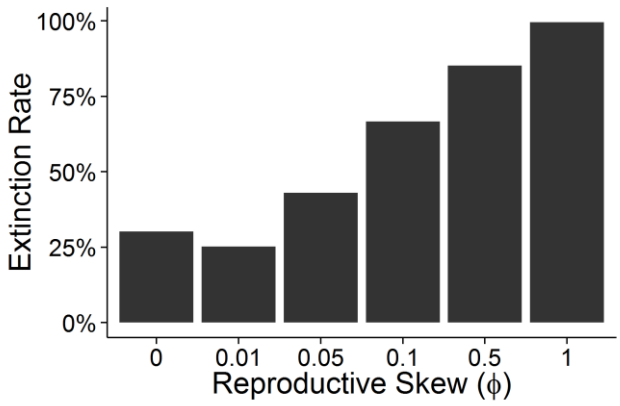
629 Second, for these same reasons, Figure 7b reveals the tendency to use social learning
630 decreases with greater reproductive skew. We return to this when we discuss the Cumulative
631 Cultural Brain Hypothesis.



(a)



(b)



(c)

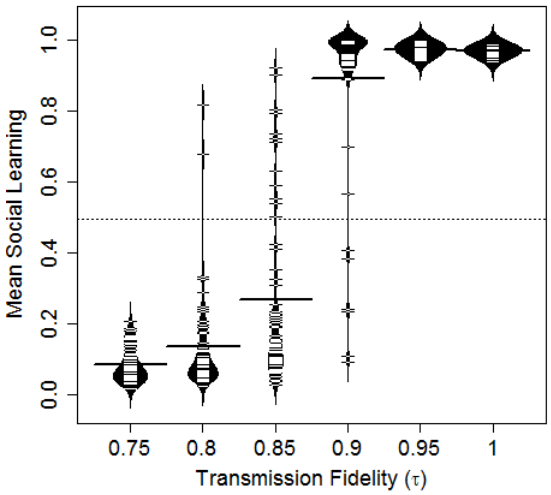
Figure 7. Bean plots showing the distribution of (a) brain size and (b) social learning means for different values of φ. The dotted horizontal line shows the global mean and the bolded horizontal lines show the group means. Bean plots show the distribution of values. (c) Plot showing the rate of extinction for different values of φ.

Empirically, these patterns are consistent with current data: brain size correlates with mating structure in both mammalian and avian lineages [6, 73]. Indeed, the relatively high rates of social learning in avian species may be due to their relatively low reproductive skews.

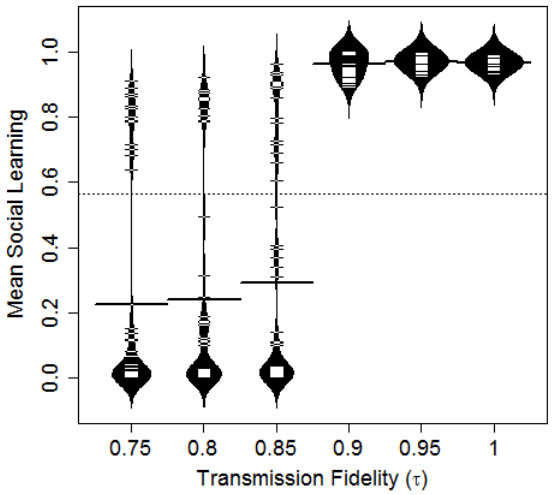
Transmission Fidelity and Asocial Learning Efficacy ($\boldsymbol{\tau}$ and $\boldsymbol{\zeta}$)

Transmission fidelity (τ) affects the degree of loss of information in the transmission of adaptive knowledge from cultural models to learners. Asocial learning efficacy (ζ) affects the efficiency with which individuals can generate new adaptive knowledge based on their own brain size. In a world of asocial learners, the parameters under which social learning is favored is narrow [recapitulating the insight from 24]. By starting in a world where the ancestral population has a lot of social learning, we gain two key insights. First, since there is little adaptive knowledge for social learners to take advantage of, we see that asocial learning is initially favored. We discuss this in detail later in the Results. Second, with an expanded range in which social learning is favored, we see how τ and ζ interact in interesting ways to affect the evolution of social learning with consequent effects on brain size, population size, etc. In Figure 8, we plot transmission fidelity against social learning for different levels of asocial learning efficacy where simulations were started with all social learners. Figure 8 shows how social learners can stand on the shoulders of effective asocial learners whose knowledge they exploit. Social learners benefit from smart ancestors.

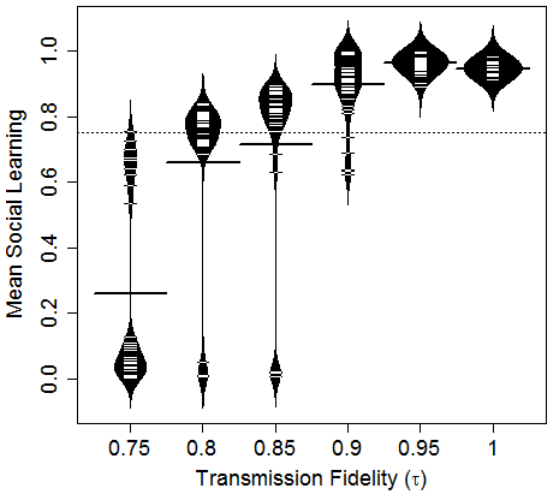
Although we treat τ and ζ as parameters in our model, we suspect that if they were allowed to evolve, they would both be pushed higher, as would reliance on social learning. And of course, larger brains that evolve via social learning will also be capable of more potent *asocial* learning since asocial learning is dependent on brain size—both in our model and in reality [see 16]. We will return to this when we discuss the Cumulative Cultural Brain Hypothesis.



$\zeta = 0.1$



$\zeta = 0.4$



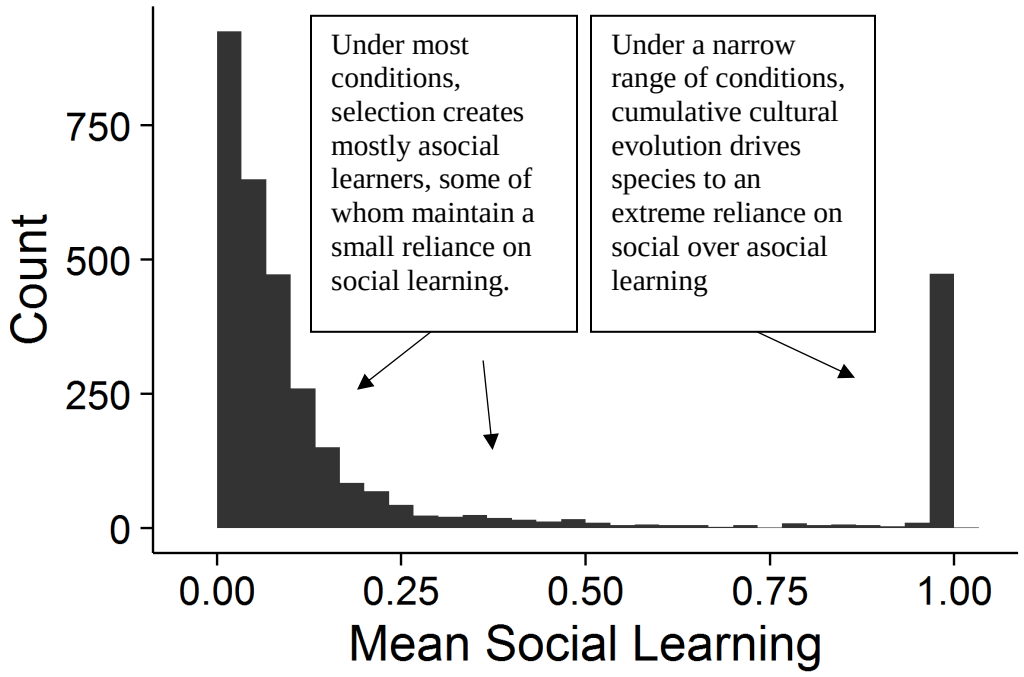
$\zeta = 0.7$

Figure 8. Bean plots showing the distribution of social learning for different values of transmission fidelity (τ) and asocial learning efficacy (ζ). The dotted horizontal line shows the global mean and the bolded horizontal lines show the group means. Bean plots show the

distribution of values. Transmission fidelity interacts with asocial learning efficacy to generate high equilibrium reliance on social learning.

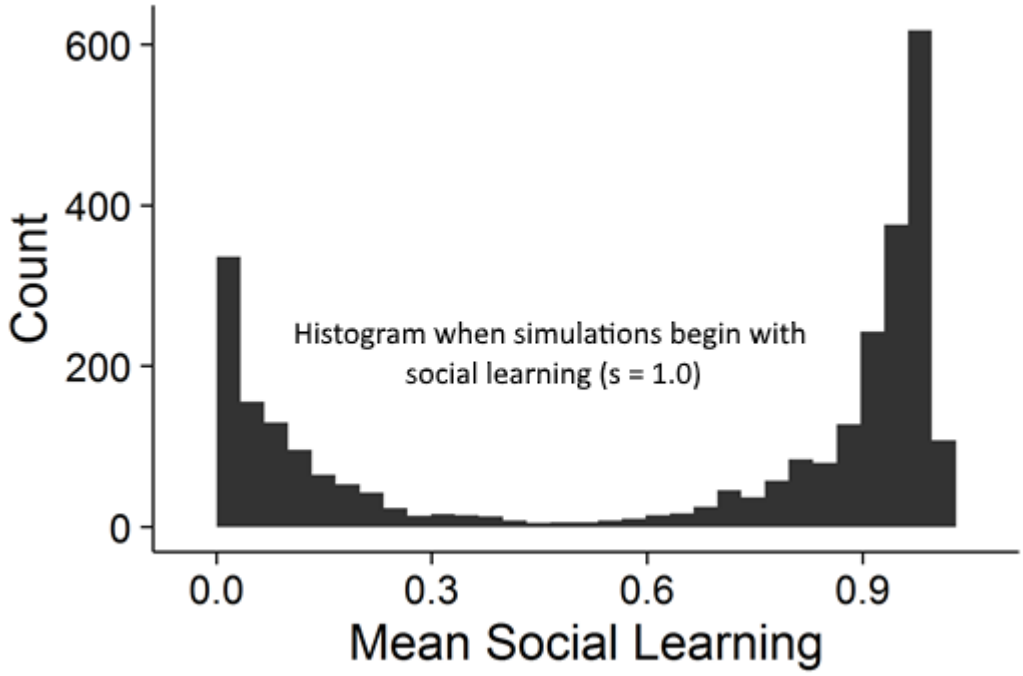
To most effectively compare our theoretical findings to the existing empirical data, we subject our simulation output to the same kinds of analyses used by researchers in the empirical literature. Of course, this comparison is qualitative: we didn't select parameter values to fit the empirical literature, but instead sought to use a wide range of plausibly realistic values, so we don't expect exact matches between the empirical correlations and our theoretical predictions. There's little doubt that some of our parameter setting never or rarely occur in the real world.

Predictions. Our range of parameters results in a range of simulated quasi-species (referred to as "species" from herein) with predicted relationships between the characteristics of these species. We have 4 key parameters in our model: Reproductive skew (mating structure; φ), transmission fidelity (τ), asocial learning efficacy (ζ), and richness of the ecology (λ). Each represents different ecological and phylogenetic constraints. The species that emerge under different combinations of these conditions can be partitioned into at least two regimes (Figure 9): species that mostly rely on (1) asocial learning or (2) social learning. A k-cluster analysis on the mean social learning value (s) for each simulation run suggests that the threshold between these regimes is approximately 50%. Note that the relative count size of the two regimes is a reflection of the range of parameters we chose rather than a reflection of the world (e.g., transmission fidelity values greater than 75%, rather than from 0% to 100%). Under some conditions, a species that mainly relies on social learning can enter into the realm of cumulative cultural evolution. The conditions that predict this transition are the basis of the CCBH. The relationships between equilibrium state variable values differ considerably between these two regimes and so we analyze them separately. The species that mostly rely on social learning include those in the realm of cumulative cultural evolution.


Under most conditions, selection creates mostly asocial learners, some of whom maintain a small reliance on social learning.
Under a narrow range of conditions, cumulative cultural evolution drives species to an extreme reliance on social over asocial learning
Count
750
500
250
0
0.00
0.25
0.50
0.75
1.00
Mean Social Learning

691

692 (a)


Histogram when simulations begin with social learning (s = 1.0)
Count
600
400
200
0
0.0
0.3
0.6
0.9
Mean Social Learning

693

(b)

Figure 9. (a) Histogram of mean social learning probability (s). Under most conditions, selection creates individuals primarily reliant on asocial learning, some of whom maintain a small reliance on social learning. Under a narrow range of conditions, cumulative cultural evolution drives species to an extreme reliance on social over asocial learning. Consistent with previous models [e.g. 24], this range of conditions expands if social learning is assumed to exist in the ancestral species; i.e., if we start the simulation with social learners. (b) Histogram of mean social learning probability (s) when simulations began with all social learners ($s = 1.0$).

To confirm that the relationships we report are not driven by cumulative cultural species (humans or hominins), we also ran a k-cluster analysis assuming 3 regimes. This analysis split species into primarily asocial learners ($s < 0.20$; e.g., beetles and buffalo), a few species with some reliance on social learning ($0.20 < s < 0.66$; e.g., capuchins and chimpanzees), and species that are almost entirely reliant on social learning ($s > 0.66$; e.g., humans, hominins, and close cousins). We then show that the relationships we find among species that mainly rely on asocial learning ($s < 0.50$) also hold among highly asocial learning species ($s < 0.20$), and relationships we find among species that mainly rely on social learning ($s > 0.50$) also hold among species with some social learning ($0.20 < s < 0.66$).

Testing Predictions. We can test our theoretically-derived qualitative predictions by comparing the species that emerge in our simulation with empirical data. Table 2 reports the relationships between the evolved characteristics of our species for each regime in our range of parameters. Below, we feature 4 key predicted relationships—(1) brain size vs. group size, (2) brain size vs. social learning, (3) brain size vs. juvenile period, and (4) group size vs. juvenile period.

717

718 **Table 1. Correlations for each regime across our entire parameter space. Correlations**
719 **between log mean brain size, log mean adaptive knowledge, log mean group size, mean**
720 **social learning, and mean juvenile period with 95% confidence intervals in brackets. The**
721 **table has been color coded from red ($r = -1$) to white ($r = 0$) to blue ($r = 1$) for ease of**
722 **comprehension. The upper table has correlations across the entire parameter space. The**
723 **lower table has primarily asocial learners ($s <.5$) in the bottom triangle and primarily**
724 **social learners ($s >.5$) in the top triangle. Following the empirical literature, social**
725 **learning is defined as the number of observed incidents of social learning. Thus, we**
726 **multiplied s by mean group size (N), and then following the empirical work, added 3, and**
727 **took the natural log [46]. The juvenile period is defined as the probability of socially**
728 **learning in a second round of learning (sv). Higher sv values should demand a longer**
729 **juvenile period.**

	Evolving characters	Brains $\log(\bar{b})$	Adaptive Knowledge $\log(\bar{a})$	Group Size $\log(\bar{N})$	Social Learning $\log(3+\overline{Ns})$	Juvenile Period $\overline{sv}$
Across entire parameter space	$\log(\bar{b})$	1				
	$\log(\bar{a})$	0.81 [.78,.82]	1			
	$\log(\bar{N})$	0.51 [.48,.53]	0.62 [.60,.64]	1		
	$\log(3+\overline{Ns})$	0.17 [.14,.20]	0.42 [.39,.45]	0.64 [.62,.66]	1	
	$\overline{sv}$	0.05 [.01,.08]	0.24 [.21,.27]	0.27 [.24,.30]	0.81 [.80,.82]	1
Primarily asocial: bottom	$\log(\bar{b})$	1	0.99 [.98,.99]	0.72 [.67,.75]	0.72 [.68,.76]	0.17 [.09,.25]
	$\log(\bar{a})$	0.78 [.77,.80]	1	0.69 [.65,.74]	0.70 [.66,.74]	0.15 [.06,.23]
	$\log(\bar{N})$	0.42 [.39,.45]	0.55 [.53,.58]	1	0.98 [.98,.98]	0.22 [.14,.30]
	$\log(3+\overline{Ns})$	-0.23 [-.26,-.19]	0.13 [.10,.17]	0.61 [.58,.63]	1	0.22 [.14,.30]
	$\overline{sv}$	-0.53 [-.56,-.51]	-0.34 [-.37,-.30]	-0.21 [-.25,-.18]	0.42 [.39,.45]	1

730

731

732 Brain Size and Group Size

733 As Table 2 shows, our model indicates that among species that *mainly rely on social*
734 *learning* (defined as $s > 0.5$), the relationship between brain size and group size is $r =$
735 0.72 [0.68,0.76]. Among species with some social learning ($0.20 < s < 0.66$), the correlation is
736 similarly, $r = 0.72$ [0.66,0.77]. In contrast, our model predicts that among taxa that rely more on
737 asocial learning, the relationship is much weaker, $r = 0.42$ [0.39,0.45]. Among highly asocial
738 learners ($s < 0.20$), the correlation is $r = 0.35$ [0.30,0.37].

739 The empirical literature has established a strong positive relationship between brain size
740 and group size in primates, but not in other taxa [32-34, 74]. In primates, the correlation between
741 relative neocortex size and group size is somewhere between $r = 0.48$ to $r = 0.61$ [52]. We
742 contrast our theoretical predictions to the empirical data in Figure 10. In support of the SBH,
743 researchers have noted that in other taxa, brain size correlates with measures of sociality or social
744 group complexity [e.g., among non-primate mammals 6, 33, 34] and with mating structure [e.g.
745 among birds; 73]. However, why group size correlates with brain size in some taxa and not others
746 remains a mystery [4]. The CBH offers an explanation, predicting that the strength of the
747 relationship between brain size and group size increases with reliance on social learning due to
748 larger groups offering a greater number of opportunities for social learning and a greater amount
749 of information to learn. Thus, for example, we should expect (and do see) a relationship between
750 brain size and group size in primates, but not ungulates or carnivores [who display less social
751 learning; 15].

752

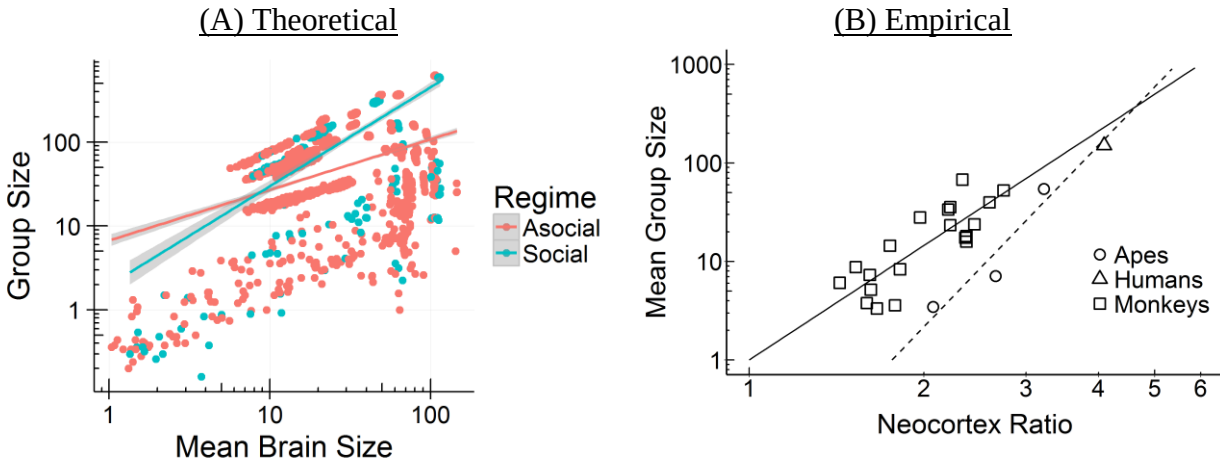


Figure 10. Brain size and group size. (A) Our model's empirical correlations between brain size and group size ($r = 0.42$ [Asocial], $r = 0.72$ [Social]). (B) Empirical correlation between brain size and group size from Barton (52) is somewhere between $r = 0.48$ to $r = 0.61$.

Notably, the algorithms in our theoretical model do not specify a direct relationship between brain size and group size or group size and brain size—these relationships just emerge. Instead, the CBH assumes that larger brains are better at storing and managing adaptive knowledge. There are two pathways to acquire that knowledge: asocial learning and social learning. Groups with higher mean adaptive knowledge are assumed to have higher carrying capacity, thus taxa more reliant on asocial learning generally also have a positive relationship between brain size and group size in our model. For taxa more reliant on social learning, larger groups also have more adaptive knowledge to exploit, raising the mean adaptive knowledge of the group and therefore the carrying capacity. Thus, our model predicts a stronger relationship between brain size and group size among taxa more reliant on social learning (compared to those more reliant on asocial learning).

768 Brain Size and Social Learning

769 Our simulations reveal a positive relationship between brain size and social learning across
770 species. Among species that primarily rely on social learning ($s > 0.5$), the relationship between
771 brain size and social learning is $r = 0.72$ [0.68,0.76]. Among species with some social learning
772 ($0.20 < s < 0.66$), the correlation is $r = 0.58$ [0.49,0.65]. However, among species that
773 primarily rely on asocial learning ($s < 0.5$), the relationship is negative, $r =$
774 -0.23 $[-0.26, -0.19]$ and similarly in strongly asocial learning species, ($s < 0.20$): $r =$
775 -0.24 $[-0.27, -0.20]$. Most asocial learning species remain small brained, but those that do
776 acquire larger brains via genetically-hardwired asocial learning do so at the expense of much
777 reliance on social learning abilities.

778 It bears emphasis that the trade-off here is between *time* or *effort* spent on asocial vs. social
779 learning, not between brain tissue allocation. If you are doing asocial learning—say running trial
780 and error to improve a tool—you can't be carefully watching others at the same time. Or,
781 alternatively, sometimes the suggested behavior delivered by asocial vs. social learning processes
782 will be contradictory, and organisms have to decide which source they will rely on. In both of these
783 senses, there's an unavoidable trade-off between social and asocial learning. However, in our
784 model, bigger brains are always better at asocial learning (when they do it), even if the selection
785 pressure that drove that brain expansion was due to the effects of social learning. That is, we
786 assume complementarity as suggested by Reader, Hager (20), and Reader and Laland (46).

787 From the empirical literature, social learning is measured by observational counts of social
788 learning events, and reveals a correlation with brain size of $r = 0.69, p < 0.001$ ($r = 0.36, p <$
789 0.05, controlling for phylogeny) for primates [46, 53]. To better match our social learning

probability, s, to the empirically available results, we assumed that simulated species with larger populations and higher s values would generate greater numbers of observational counts (linearly). Thus, we multiplied s by mean group size (N), and then following the existing empirical approach, added 3, and took the natural log [46]. A similar relationship has been shown for birds using indirect measures of opportunities for social learning [e.g. number of caretakers; 23]. Figure 11 contrasts our predicted relationship with the empirical literature.

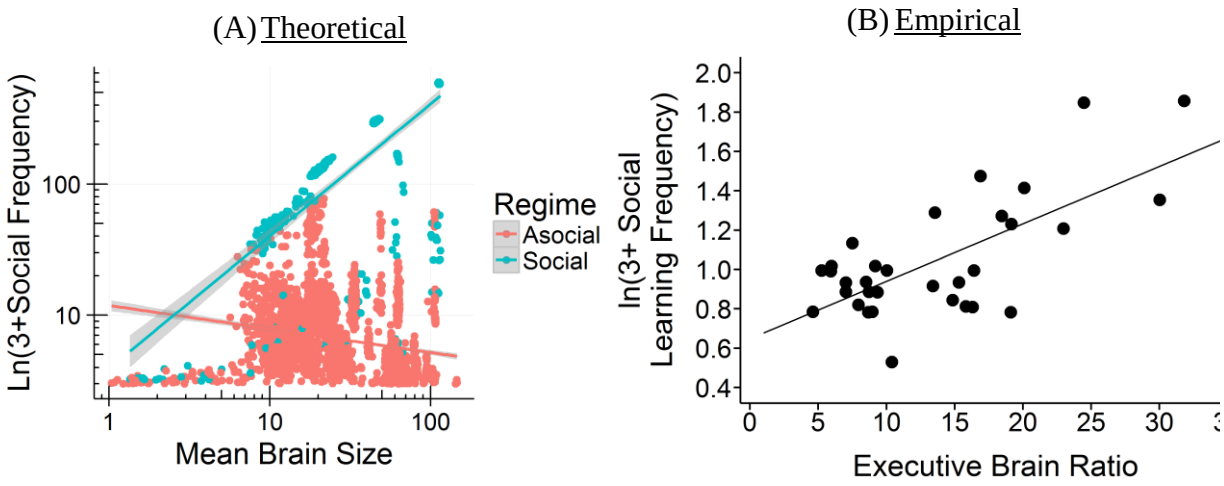


Figure 11. Brain size and social learning. (A) Our model's empirical correlations between brain size and incidences of social learning ($r = -0.23$ [Asocial], $r = 0.72$ [Social]). (B) Empirical correlation between brain size and incidences of social learning among primates from Reader and Laland (46) is $r = 0.69$ ($r = 0.36$ controlling for phylogeny). A similar relationship has been shown for birds using indirect measures of opportunities for social learning [e.g. number of caretakers; 23].

Brain Size and Juvenile Period

Our simulation does not explicitly model the length of the extended juvenile period, but does include 2 periods of learning. In the first period, individuals can learn socially from their

genetic parent or asocially by themselves. In the second period, individuals with a low s value are likely to update their knowledge asocially, while those with higher s values only updated their knowledge obliquely based on their v value; individuals had a $1-s$ probability of updating asocially, sv probability of updating socially and an $s-sv$ probability of doing no further learning. Thus, sv represents an extended juvenile period in which learners could use payoff-biased oblique transmission to update their knowledge. Larger sv values should demand a longer juvenile period.

Our model indicates that among species that mainly rely on social learning ($s > 0.5$), the relationship between brain size and the length of the extended juvenile period is $r = 0.17$ [0.09,0.25]. This positive relationship only occurs when we include highly social learners ($s > 0.66$). The relationship between brain size and the length of an extended juvenile period disappears or is negative among species with only a moderate reliance on social learning ($0.20 < s < 0.66$), $r = -0.08$ $[-0.20, 0.04]$, more reliant on asocial learning ($s < 0.5$), $r = -0.53$ $[-0.56, -0.51]$, or are highly asocial ($s < 0.20$), $r = -0.59$ $[-0.61, -0.56]$. Thus, we argue that an extended juvenile period evolves to support more opportunities to engage in social learning.

Our *extended* juvenile period most closely represents an adolescent period (the period from sexual maturity to sexual reproduction), where additional biased oblique social learning occurs. Adolescence is rare, occurring in humans, possibly elephants [71] and some orca [72], and some members of cooperative breeding species [75]. Nonetheless, positive relationships between brain size and the length of the *juvenile* period (weaning age to sexual reproduction) have been shown directly in primates [54, 56, 57] and indirectly via age to sexual maturity in a variety of taxa [55].

The correlation for primates is $r = 0.61, p = 0.037$ [56]. Though the comparison is imperfect, we show the relationship between brain size and length of the extended juvenile period side by side with the relationship between brain size and the juvenile period in primates in Figure 12 below.

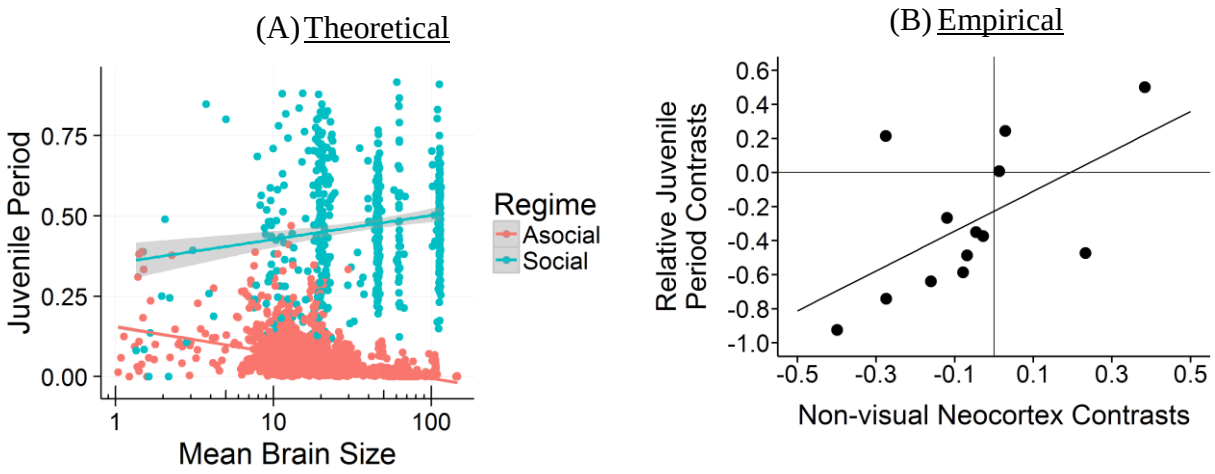


Figure 12. Brain size and the juvenile period. (A) Our model's empirical correlations between brain size and the length of the extended juvenile period ($r = -0.53$ [Asocial], $r = 0.17$ [Social]). (B) Empirical correlation between brain size and juvenile period among primate species from Joffe (56) is $r = 0.61$.

Group Size and Juvenile Period

Since an extended juvenile period primarily evolves in the presence of large amounts of adaptive knowledge that requires more opportunities for social learning, we should also expect to see a positive relationship between group size and the juvenile period among highly social learners. Indeed, our model indicates that among species that mainly rely on social learning ($s > 0.5$), the relationship between group size and the length of the juvenile period is $r = 0.22$ [0.14,0.30]. As with the relationship between brain size and the length of the extended juvenile period (and for related reasons), this positive relationship only occurs when we include highly social learners ($s >$ 0.66). The relationship between brain size and the length of an extended juvenile period disappears

or is negative among species with only a moderate reliance on social learning ($0.20 < s < 0.66$), $r = -0.05$ $[-0.16, 0.07]$, mainly rely on asocial learning ($s < 0.5$), $r = -0.21$ $[-0.25, -0.18]$, or are highly reliant on asocial learning ($s < 0.20$), $r = -0.12$ $[-0.16, -0.08]$. For highly social learning species, this positive relationship is an indirect consequence of social learners having access to more knowledge in larger groups, creating a stronger selection pressure for a longer juvenile period in which to take advantage of this knowledge. This, in turn, raises the average adaptive knowledge of the group, allowing for larger groups. Empirically, in primates, the relationship between absolute juvenile period length (we were unable to find the weaning age to sexual maturity measure; sexual maturity to sexual reproduction is non-existent) and mean group size is $r = 0.57, p = 0.007$ [56]. In Figure 13 below, we contrast our predictions against the empirical results. Joffe (56) did not provide a comparison plot, but we have generated one from his data.

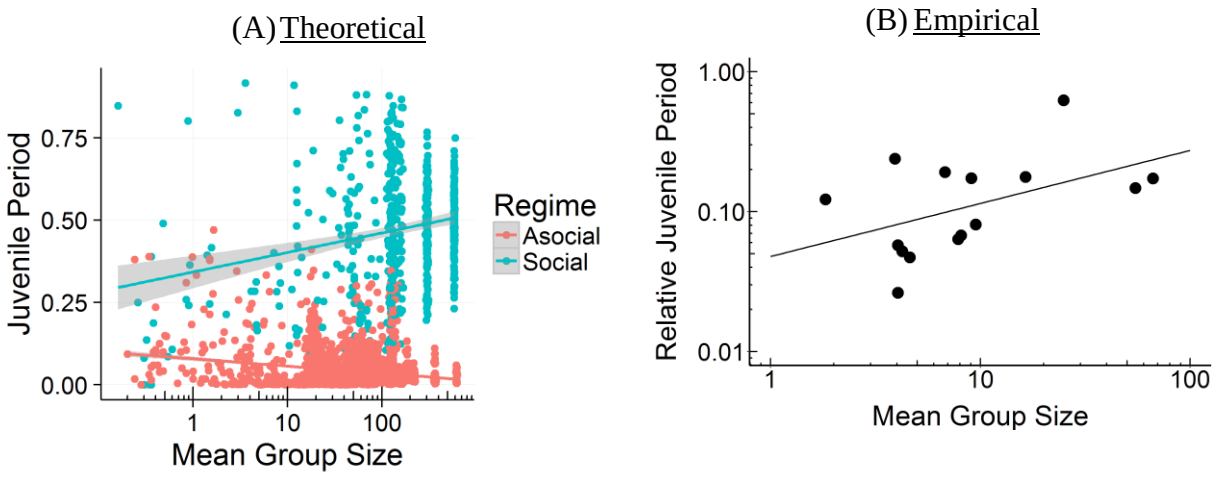


Figure 13. Group size and the juvenile period. (A) Our model's empirical correlations between group size and the length of the juvenile period ($r = -0.21$ [Asocial], $r = 0.22$

[Social]). (B) Empirical correlation between group size and the length of the juvenile period among primates from Joffe (56) is $r = 0.57$.

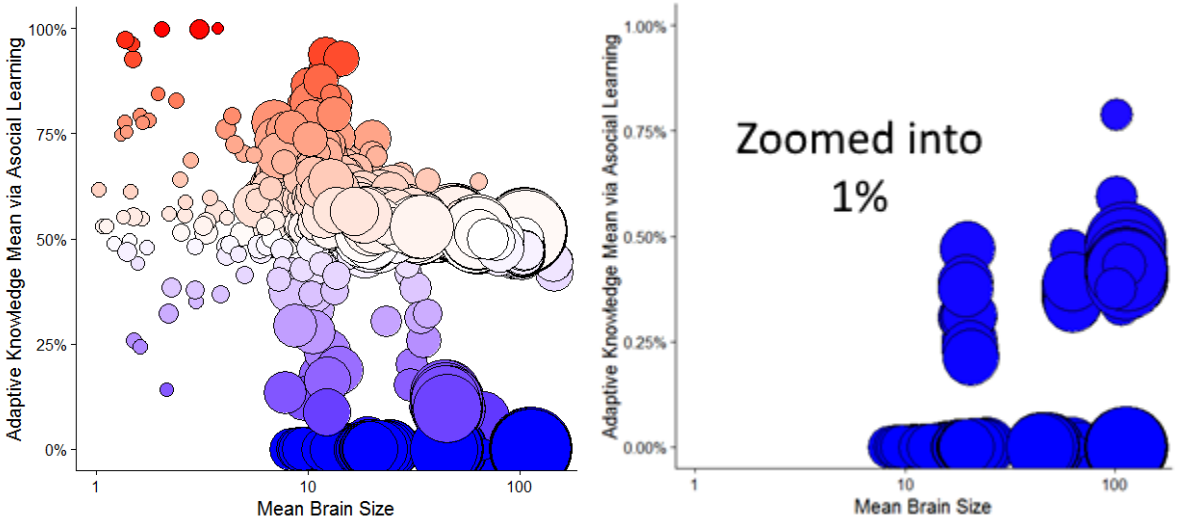
The Cumulative Cultural Brain Hypothesis

Beyond the hypothesis that social learning, brain size, adaptive knowledge, and group size may have coevolved to create the patterns found in the empirical literature, we are also interested in the conditions under which these variables might interact synergistically to create highly social species with large brains and substantial accumulations of adaptive knowledge (humans). To assess when an accumulation of adaptive knowledge becomes cumulative cultural evolution, we apply a standard definition of cumulative cultural products as being those products that a single individual could not invent by themselves in their lifetime. To calculate this for our species, we ask what the probability is that an individual with the average brain size of the species would invent the mean level of adaptive knowledge in that species via asocial learning.

Formalization of Cumulative Culture. The probability of an individual i in deme j acquiring the mean deme adaptive knowledge A_j through asocial learning is given by Equation 8. Asocial learners draw their adaptive knowledge value from a normal distribution with mean of their brain size scaled by ζ. Thus the probability of acquiring $a_{ij} \geq A_j$ is the integral from this mean value or greater over the asocial learning distribution. Note that this gives the probability of an individual acquiring that level of adaptive knowledge. The probability that the mean adaptive knowledge of the deme is reached through asocial learning is this probability to the power of the number of individuals in the deme ($P\left[a_{ij} \geq A_j\right]^{N_j}$)—a slim chance indeed.

$$P\left[a_{ij} \geq A_j\right] = \int_{A_j}^{\infty} N\left(\zeta b_{ij}, \sigma_a \zeta b_{ij}\right) \tag{8}$$

We set a low, albeit arbitrary, threshold where the probability of any individual acquiring this level of adaptive knowledge through asocial learning is less than 0.1%. At this level, the probability that an entire population would develop that level of adaptive knowledge through asocial learning is 0.001^{N_j}, i.e., exceedingly unlikely. Thus, mean levels of adaptive knowledge that are so exceedingly unlikely to have been acquired through asocial learning can be attributed to cumulative cultural evolution. In Figure 14 below, we plot brain size against the probability of acquiring that amount of information.



(a) (b)

Figure 14. Cumulative Culture and Brain Size. Circle size indicates the mean population size. More red indicates high probability of acquiring knowledge through asocial learning and more blue indicates a low probability. The darkest blue circles in the bottom right are the simulations that cross the threshold into the cumulative cultural realm. (a) Log mean brain size against the probability of acquiring the mean adaptive knowledge in the group via asocial learning. (b) Here we show the same data zoomed in-between 0 and 1%.

Next, we can look at what parameters increase and decrease the probability of entering into the realm of cumulative cultural evolution, on the bottom right corner of Figure 14a: those large-brained species with a lot of adaptive knowledge, which they were unlikely to acquire without cumulative cultural evolution.

Transmission Fidelity Drives Larger Brains. The simulation predicts that transmission fidelity is the key to entering into the realm of cumulative cultural evolution. When the model begins with widespread social learning, we see a threshold effect, where for very high fidelity transmission ($\tau > 0.85$), social learning and large brains evolve under a wide range of parameters. However, when we begin with primarily asocial learners (more plausible), this threshold increases to nearly 100% (see Figure 16). The degree of these results may be exaggerated by our "stacking the deck" against social learning, but the overall results are consistent with previous models that argue that transmission fidelity is the key to cumulative cultural evolution [29]. And also with models that show that there is a fitness valley that needs to be crossed to enter into the realm of cumulative culture and reliance on social learning [24]. When social learning is already present in the population, species can enter the realm of cumulative cultural evolution under a wider range of parameters—that is, the more pre-existing social learning exists, the shallower the fitness valley that needs crossing.

Embedded in τ, and eventually oblique learning and learning bias, are cognitive abilities like theory of mind, the ability to recognize, distinguish, and imitate potential models, but also teaching and social tolerance. We suspect that if we endogenized τ, either cultural or genetic evolution would favour higher values under these conditions.

Reprodutive Skew or Individual vs Between-Group Selection Matters. As discussed, lower reproductive skew consistent with "monogamish" or human cooperative/communal mating

structures [62] and consistent with less individual selection and increased between-group selection are more likely to lead to social learning and therefore to cumulative cultural evolution. These results are consistent with recent papers explicitly modeling these levels of selection [38, 39]. Too strong an individual selection pressure leads to bigger brains via asocial learning—bigger mutant brains can't be filled via social learning, since cultural information is capped by brain size—but these populations often go extinct, even when we start with fully developed social learning. We graph the probability of entering into the realm of cumulative cultural evolution for different values of φ in Figure 16a. We see a Goldilocks' zone around $\varphi = 0.01$ (significantly higher than both $\varphi = 0$ and $\varphi = 0.05$), regardless of the starting conditions (though as previously discussed, the parameter range leading to cumulative culture increases if social learning is common). As reproductive skew or individual selection increases, asocial learning is favored. Thus, entering the realm of cumulative cultural evolution is less likely.

Smart Ancestors and Rich Ecologies. As discussed in (1), we find that an interaction between transmission fidelity τ and individual learning ζ fuels the autocatalytic take-off. If ζ is too high, individual learning is too efficient and social learning struggles to take flight, except at very high rates of transmission fidelity or if social learning is already present. But if ζ is too low, even if social learning out-competes individual learning, populations have smaller brains and less adaptive knowledge compared to when social learning out-competes more effective individual learning. These results suggest that social learners stand on the shoulders of effective asocial learners. That is, when social learning can initially exploit the adaptive knowledge developed by more effective individual learning, social learning results in larger brains. The Cumulative Cultural Brain Hypothesis predicts innovative ancestors—perhaps like the kind of individual innovativeness we see in chimpanzees [20].

Finally, environments have to be sufficiently rich (λ) and have good survival returns to adaptive knowledge to open the door to the regime of cumulative cultural evolution. Brains are costly, but this cost can be offset by more adaptive knowledge. The degree of mitigation is determined by λ. We find that higher λ values allow for the evolution of larger brains. Basically, you need to be in an environment where adaptive knowledge pays off well enough to pay for those costly brains.

One interesting, but speculative possibility that links these two parameters is that as the East African cradle of human evolution cooled and forests became savannah[76], our ancestors may have faced an increased selection for smaller brains helping to trigger the transition from asocial to social learning. That is, the forests were a richer ecology with higher λ, allowing for large-brained ancestors who could pay for their large brains through asocial smarts. As the forest thinned into savannah, the ecology became tougher and λ decreased, social learning may have provided a cheaper alternative to acquiring this knowledge and gaining more in order to maintain large brains in a calorie-poorer and less forgiving environment. Though we might infer such a scenario from our model, we would need to adjust these parameters within the model in order to test this hypothesis.

Why some social learning is common but cumulative cultural evolution is rare

In addition to our main simulations that began with asocial learners, we also ran a set of simulations that began with social learners. Although social learning is widespread in the animal kingdom [22] and the most realistic starting conditions are somewhere in-between these two extremes (no social learning and complete social learning), these realistic conditions are likely closer to no social learning than complete social learning. Nonetheless, running our simulations

beginning with social learning provides an upper bound on our predicted patterns and also offers additional insights.

There are two key insights. The first is that social learning is maladaptive in a world with little knowledge (Figure 15). With little knowledge for social learners to exploit, asocial learners quickly invade. However, since some social learning is present, once sufficient knowledge has been generated, social learning is again at an advantage, with additional innovations generated in the process of social learning [16]. The second key insight is closely related: consistent with previous models [24], the presence of social learning expands the range of parameters in which cumulative culture is adaptive. Figure 9b shows a greater number of species with social learning (compared to Figure 9a). Figure 16a reveals that more monogamish societies, or at least societies with a reduced reproductive skew or reduced individual selection are more likely to enter the realm of cumulative cultural evolution. Figure 16b reveals that cumulative cultural evolution is more likely to evolve when transmission fidelity is higher. Both Figures 16a and 16b reveal that the range of parameters that lead to the realm of cumulative cultural evolution expands if more social learning is present in the ancestral state.

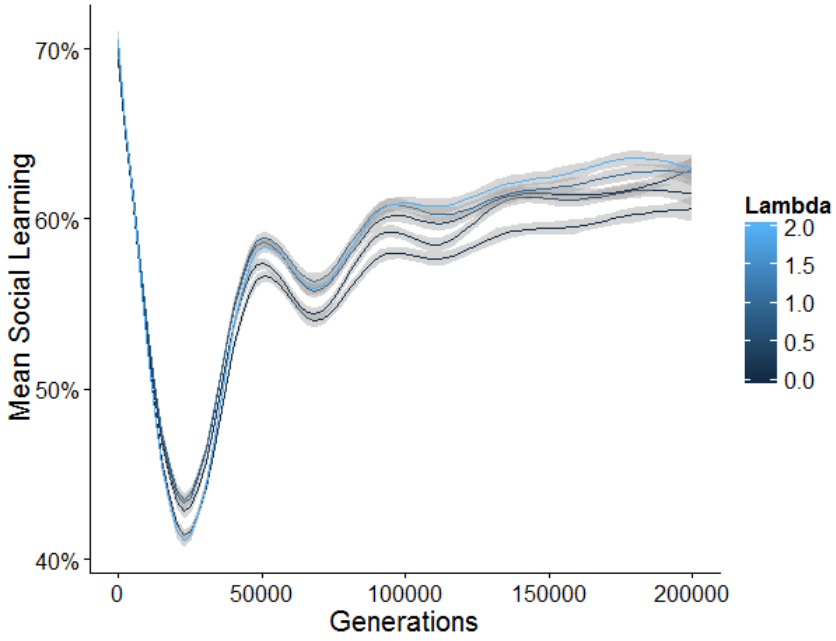
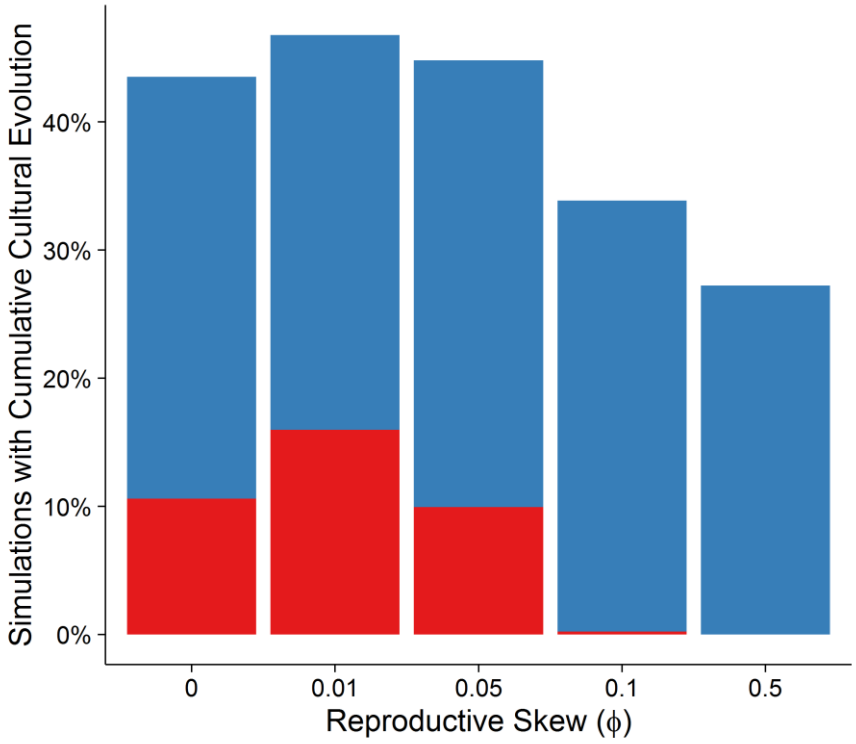
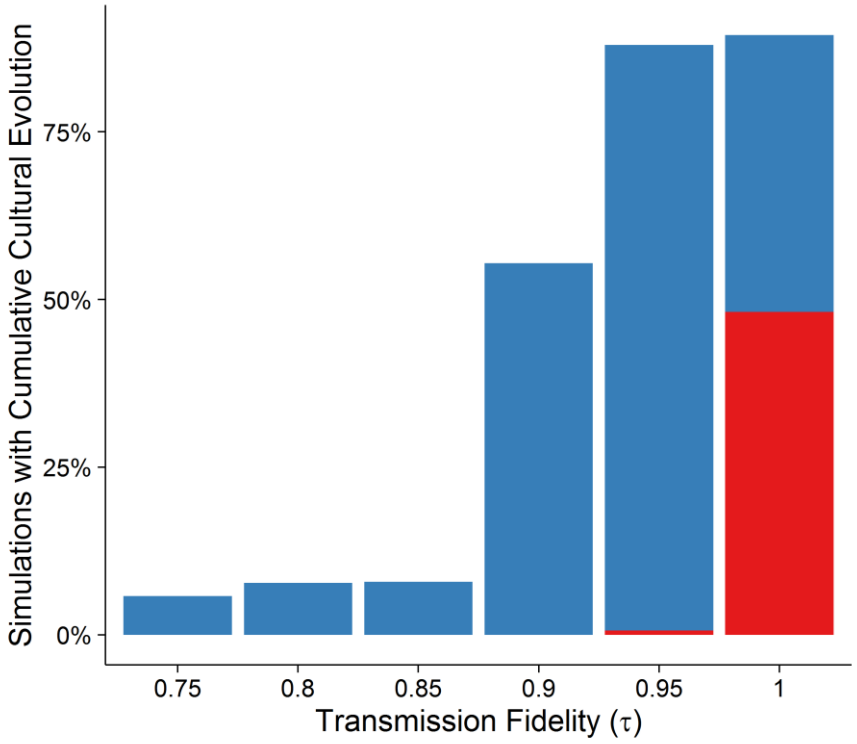


Figure 15. Social learning over generations starting with $s = 1.0$. Social learning is maladaptive in the absence of adaptive knowledge. Asocial learners quickly invade. It is only when asocial learners have generated sufficient adaptive knowledge that social learners again have an advantage. Since we know that at least two regimes reliably emerge, mean social learning in these plots represents the relative number of conditions in which social and asocial learners emerge rather than a value of social learning characteristic of the world.


Simulations with Cumulative Cultural Evolution
40%
30%
20%
10%
0%
0
0.01
0.05
0.1
0.5
Reproductive Skew (ϕ)

984

985 (a)



(b)

Figure 16. Percent of simulations in which cumulative cultural evolution evolves. Blue simulations are those that began with $s = 1.0$ and red simulations are those that began with $s = 0.0$. (a) across different values of reproductive skew (ϕ) and (b) across different values of transmission fidelity (τ).

Discussion

In this discussion section we (1) summarize our key findings, (2) review these findings in the context of the cultural/general intelligence hypotheses and related work, and (3) discuss limitations of this work and ongoing inquiries.

Summary of Key Findings

Our model provides a potential evolutionary mechanism that can explain a variety of empirical patterns involving relationships between brain size, group size, innovation, social learning, mating structures, and developmental trajectory, as well as brain evolution differences among species. It can also illuminate the different rates of evolution and overall brain size that have been found in different taxa and help explain why brain size correlates with group size in some taxa, but not others. In contrast to competing explanations, the key message of the Cultural Brain Hypothesis (CBH) is that brains are primarily for the acquisition, storage and management of adaptive knowledge and that this adaptive knowledge can be acquired via asocial or social learning. Social learners flourish in an environment filled with knowledge (such as those found in larger groups and those that descend from smarter ancestors), whereas asocial learners flourish in environments where knowledge is socially scarce, or expensive but obtainable through individual efforts. The correlations that have been found in the empirical literature between brain size, group size, social learning, the juvenile period, and adaptive knowledge arise as an indirect result of these processes.

The Cumulative Cultural Brain Hypothesis posits that these very same processes can, under very specific circumstances, lead to the realm of cumulative cultural evolution. These circumstances include when transmission fidelity is sufficiently high, reproductive skew is in a Goldilocks' zone close to monogamy (or equally, there is some, but not too much individual-level selection), effective asocial learning has already evolved, and the ecology offers sufficient rewards for adaptive knowledge. In making these predictions, the Cultural Brain Hypothesis and Cumulative Cultural Brain Hypothesis tie together several lines of empirical and theoretical research.

Related Work

Under the broad rubric of the Social Brain or Social Intelligence Hypothesis, different researchers have highlighted different underlying evolutionary mechanisms [35-38, 40]. These models have had differing levels of success in accounting for empirical phenomena, but they highlight the need to be specific in identifying the driving processes that underlie brain evolution in general, and the human brain specifically. From the perspective of the CBH, these models have been limited in their success, because they only tell part of the story. Our results suggest that the CBH can account for all the empirical relationships emphasized by the Social Brain Hypotheses, plus other empirical patterns not tackled by the SBH. Moreover, our approach specifies a clear 'take-off' mechanism for human evolution that can account for our oversized crania, heavy reliance on social learning with sophisticated forms of oblique transmission (and possibly the emergence of adolescence as a human life history stage), and the empirically-established relationship between group size and toolkit size/complexity [77]—as well as, of course, our species' extreme reliance on cumulative culture for survival [19].

Our results echo some of the predictions of models of learning and levels of competition. In particular, an early paper by Gavrilets and Vose [40] pitched as a model of Machievellian intelligence might equally be viewed as a model of culture, showing similar co-evolution of brain size, adaptive knowledge and learning ability. A more recent paper by Gavrilets [38] modeled socio-cognitive competencies in competition between groups, between individuals, and against the environment. This model showed how socio-cognitive competencies were enhanced with weaker individual-level selection, which is echoed in the CBH predictions. Finally, a recent paper by González-Forero and Gardner [39] model the energy tradeoff between brains, bodies, and reproduction under different challenges and costs. This energy model takes a different approach

to how the variable and parameters are specified, particularly in tracking in ratios of brain size and energy extraction efficiency, making it difficult to directly compare to the CBH and CCBH. While the mapping is not perfect, these are potentially complementary models, particularly in the overall result that humans emerge where competition is 60% ecological, 30% cooperative, and 10% between groups with little individual-level competition, reflecting the importance of a high λ and low φ in our model. The authors conclude by noting how their model may intersect with a model of culture like the CBH in how social learning and life history interact with ecological factors and the relationship between adaptive knowledge and survival.

Our simulation's predictions are consistent with other theoretical work on cultural evolution and culture-gene coevolution. For example, several researchers have argued for the causal effect of sociality on both the complexity and quantity of adaptive knowledge [78, 79]. Similarly, several researchers have argued for the importance of high fidelity transmission for the rise of cumulative cultural evolution [29, 48, 80].

Cultural variation is common among many animals (e.g., rats, pigeons, chimpanzees, and octopuses), but cumulative cultural evolution is rare [24, 81]. Boyd and Richerson [24] have argued that although learning mechanisms, such as local enhancement (often classified as a type of social learning), can maintain cultural variation, observational learning is required for cumulative cultural evolution. Moreover, the fitness valley between culture and cumulative culture grows larger as social learning becomes rarer. Our model supports both arguments by showing that only high fidelity social learning gives rise to cumulative cultural evolution and that the parameter range to enter this realm expands if social learning is more common (see Figure 16). In our model, cumulative cultural evolution exerts a selection pressure for larger brains that, in turn, allows more culture to accumulate. Prior research has identified many mechanisms, such as

teaching, imitation, and theory of mind, underlying high fidelity transmission and cumulative cultural evolution [18, 28, 82]. Our model reveals that in general, social learning leads to more adaptive knowledge and larger brain sizes, but shows that asocial learning can also lead to increased brain size. Further, our model indicates that asocial learning may provide a foundation for the evolution of larger-brained social learners. These findings are consistent with Reader et al. [20], who argue for a primate general intelligence that may be a precursor to cultural intelligence and also correlates with absolute brain volume. And, though more speculative, key mutations, such as the recently discovered NOTCH2NL genes [83, 84], may have allowed for the transition from smart asocial learners to larger brained social learners as specified in the narrow pathway of the CCBH.

The CBH is consistent with much existing work on comparative cognition across diverse taxa. For example, in a study of 36 species across many taxa, MacLean et al. [85] show that brain size correlates with the ability to monitor food locations when the food was moved by experimenters and to avoid a transparent barrier to acquire snacks, using previously acquired knowledge. The authors also show that brain size predicts dietary breadth, which was also an independent predictor of performance on these tasks. Brain size did not predict group size across all these species (some of whom relied heavily on asocial learning). This alternative pathway of asocial learning is consistent with emerging evidence from other taxa. For example, in mammalian carnivores brain size predicts greater problem solving ability, but not necessarily social cognition [86, 87]. These results are precisely what one would expect based on the Cultural Brain Hypothesis; brains have primarily evolved to acquire, store and manage adaptive knowledge that can be acquired socially or asocially (or via both). The Cultural Brain Hypothesis predicts a strong

relationship between brain size and group size among social learning species, but a weaker or non-existent relationship among species that rely heavily on asocial learning.

Our simulation results are also consistent with empirical data for relationships between brain size, sociality, culture, and life history among extant primates [e.g. 88] and even cetaceans [34], but suggest a different pathway for humans. In our species, the need to socially acquire, store, and organize an ever expanding body of cultural know-how resulted in a runaway coevolution of brains, learning, sociality and life history. Of course, this hypothesis should be kept separate from the CBH: at the point of the human take-off, brain size may have already been pushed up by the coordination demands of large groups, Machiavellian competition, or asocial learning opportunities [19]. For example, Machiavellian competition may have elevated mentalizing abilities in our primate ancestors that were later high-jacked, or re-purposed, by selective pressure associated with the CCBH to improve social learning by raising transmission fidelity, thereby creating cumulative cultural evolution. Thus, the CBH and CCBH should be evaluated independently.

Note that in understanding these results, it is worth remembering that our model assumes a relationship between brain size and adaptive knowledge capacity, but not adaptive knowledge; similarly between adaptive knowledge and carrying capacity, but not population size; and between brain size and decreased survival and adaptive knowledge and increased survival. These tradeoffs and co-evolutionary dynamics help us understand why we see stronger or weaker relationships between social and asocial species

Synthesis and Naming. These ideas, which have been developed concurrently by researchers in different fields, are sufficiently new such that naming and labeling conventions have not yet converged. We use Cultural Brain Hypothesis and the Cumulative Cultural Brain Hypothesis for

the ideas embodied in our formal model. We nevertheless emphasize that we are building directly on a wide variety of prior work that has used various naming conventions, including The Cultural Intelligence Hypothesis [21] and the Vygotskian Intelligence Hypothesis [89]. And, of course, Humphrey (41) originally described the importance of social learning in his paper on the social functions of intellect, though subsequent work has shifted the emphasis away from social learning and toward both Machiavellian strategizing and the management of social relationships. Whiten and Van Schaik (21) first used the term "Cultural Intelligence Hypothesis" to argue that culture may have driven the evolution of brain size in non-human great apes. Later, Herrmann, Call (26) used the same term to argue that humans have a suite of cognitive abilities that have allowed for the acquisition of culture. Supporting data for both uses of the term are consistent with the CBH and the CCBH [for a rich set of data and analyses, see 20]. We used two new terms not to neologize, but because though our approach is clearly related to these other efforts, our approach contains novel elements and distinctions not clarified or formalized in earlier formulations.

Simplifications, Extensions and Future work

Note that our model seeks to (1) show why brain size, adaptive knowledge, social learning, group size, and lifespan are intercorrelated across the animal kingdom (CBH) and (2) how the very same processes that lead to these interconnections, can, under some specific circumstances, lead to the realm of cumulative cultural evolution—the uniquely human pathway. Within the realm of cumulative culture, the dynamics change in ways that are not captured by this model. For example, in order to sustain ever-growing levels of cultural complexity, cultures can generate ways to increase sociality and transmission fidelity. With sufficiently complex culture, mechanisms may evolve to more efficiently share the fruits of rare innovations, allowing for increases in cultural variance that may be individually costly. Moreover, cumulative culture, once acquired, can

increase an individual efficacy in subsequent asocial learning [for a discussion of these ideas, see 16].

In developing the simulation, we formalized the minimal set of assumptions and parameters that capture the logic of the CBH and CCBH. There are a number of extensions, variations, and additional parameters that would improve our understanding of the evolution of brain size.

There were several assumptions that simplified our model, making it more computationally tractable. Future models may address some of these shortcomings and explore additional parameters. One such improvement is to explicitly track different cultural traits with different cognitive costs and fitness payoffs. By doing this, we could better explore the benefits to migration and cultural recombination. We would also like to more fully explore the impact of the relationship between adaptive knowledge and carrying capacity. Currently, the richness of the ecology only affects individual survival based on paying the calorie cost of costly brains, but the richness of the ecology also affects the carrying capacity of the population with consequent effects for the dynamics between brain size, adaptive knowledge and population size.

Another previously mentioned future improvement is the endogenization of transmission fidelity (τ) and reproductive skew (φ). These parameters are themselves subject to genetic and cultural evolutionary processes and thus ought to be modeled as endogenous variables. In our model, we can discuss the effect of different evolutionary outcomes or values of transmission fidelity and reproductive skew, but not their evolution.

Two or three regimes emerged in our models based on different ecological and phylogenetic constraints. In a future model, we plan to explore the adaptive dynamics of these different regimes, exploring the invasion fitness of the different equilibrium states discovered in

our model. These models will help us better understand the evolutionary dynamics that may have occurred when different previously geographically separated hominin species encountered each other (e.g., the European encounter between modern humans and their larger-brained Neanderthal cousins).

The key improvements that we are eager to explore could be summarized as: (1) endogenizing the evolution of transmission fidelity and reproductive skew, (2) explicitly tracking different cultural traits with different cognitive costs and fitness payoffs, and (3) more thoroughly exploring the brain shrinkage that occurs during the transition from reliance on asocial learning to reliance on social learning[1]. These results hint that the process underlying the Cultural Brain Hypothesis and Cumulative Cultural Brain Hypothesis may also help explain evidence suggesting that human brains have been shrinking in the last 10,000 to 20,000 years [90]. Although this shrinkage in brain size corresponds to shrinking in body size, it may be evidence that our species is not at equilibrium.

Acknowledgements

The computational model was enabled in part by support provided by Westgrid and Compute Canada. J.H. acknowledges support from the Canadian Institute for Advanced Research.

References

1. Bailey DH, Geary DC. Hominid brain evolution. Human Nature. 2009;20(1):67-79.

[1] This brain shrinkage (see Figure 15) occurs as social learners invade by "stealing" the knowledge of the asocial learners, without having to figure it out for themselves. Once the population is mostly made up of social learners, brain size begins to increase again.

1173 2. Schoenemann PT. Evolution of the size and functional areas of the human brain. Annual
1174 Review of Anthropology. 2006;35:379-406.
1175 3. Striedter GF. Principles of brain evolution. Sunderland, MA: Sinauer Associates; 2005.
1176 4. Dunbar RI, Shultz S. Evolution in the social brain. Science. 2007;317(5843):1344-7.
1177 5. Roth G, Dicke U. Evolution of the brain and intelligence. Trends in cognitive sciences.
1178 2005;9(5):250-7.
1179 6. Shultz S, Dunbar R. Encephalization is not a universal macroevolutionary phenomenon in
1180 mammals but is associated with sociality. Proceedings of the National Academy of Sciences.
1181 2010;107(50):21582-6.
1182 7. Balanoff AM, Bever GS, Rowe TB, Norell MA. Evolutionary origins of the avian brain.
1183 Nature. 2013;501(7465):93-6.
1184 8. Aiello LC, Wheeler P. The expensive-tissue hypothesis: the brain and the digestive system
1185 in human and primate evolution. Current anthropology. 1995;36(2):199-221.
1186 9. Lieberman D. The evolution of the human head: Harvard University Press; 2011.
1187 10. Elia M. Organ and Tissue Contribution to Metabolic Weight. In: Kinney JM, Tucker HN,
1188 editors. Energy Metabolism: Tissue Determinants and Cellular Corollaries. New York, NY: Raven
1189 Press; 1999.
1190 11. Kotrschal A, Rogell B, Bundsen A, Svensson B, Zajitschek S, Brännström I, et al. Artificial
1191 selection on relative brain size in the guppy reveals costs and benefits of evolving a larger brain.
1192 Current Biology. 2013;23(2):168-71.
1193 12. Foley RA, Lee PC, Widdowson E, Knight C, Jonxis J. Ecology and energetics of
1194 encephalization in hominid evolution [and discussion]. Philosophical Transactions of the Royal
1195 Society B: Biological Sciences. 1991;334(1270):223-32.

1196 13. Isler K, Van Schaik CP. Metabolic costs of brain size evolution. Biology Letters.
1197 2006;2(4):557-60.
1198 14. Dunbar RI. The social brain: mind, language, and society in evolutionary perspective.
1199 Annual Review of Anthropology. 2003:163-81.
1200 15. van Schaik CP, Burkart JM. Social learning and evolution: the cultural intelligence
1201 hypothesis. Philosophical Transactions of the Royal Society B: Biological Sciences.
1202 2011;366(1567):1008-16.
1203 16. Muthukrishna M, Henrich J. Innovation in the collective brain. Philosophical Transactions
1204 of the Royal Society of London B: Biological Sciences. 2016;371(1690). doi:
1205 10.1098/rstb.2015.0192.
1206 17. Muthukrishna M. The Cultural Brain Hypothesis and the Transmission and Evolution of
1207 Culture. Vancouver, BC: University of British Columbia; 2015.
1208 18. Heyes C. Grist and mills: on the cultural origins of cultural learning. Philosophical
1209 Transactions of the Royal Society B: Biological Sciences. 2012;367(1599):2181-91.
1210 19. Henrich J. The secret of our success: How culture is driving human evolution,
1211 domesticating our species, and making us smarter. Princeton, NJ: Princeton University Press;
1212 2016.
1213 20. Reader SM, Hager Y, Laland KN. The evolution of primate general and cultural
1214 intelligence. Philosophical Transactions of the Royal Society B: Biological Sciences.
1215 2011;366(1567):1017-27.
1216 21. Whiten A, Van Schaik CP. The evolution of animal ‘cultures’ and social intelligence.
1217 Philosophical Transactions of the Royal Society B: Biological Sciences. 2007;362(1480):603-20.

1218 22. Hoppitt W, Laland KN. Social Learning: An Introduction to Mechanisms, Methods, and
1219 Models. Princeton, NJ: Princeton University Press; 2013.
1220 23. van Schaik CP, Isler K, Burkart JM. Explaining brain size variation: from social to cultural
1221 brain. Trends in cognitive sciences. 2012;16(5):277-84.
1222 24. Boyd R, Richerson PJ, editors. Why culture is common, but cultural evolution is rare.
1223 Proceedings of the British Academy; 1996: Oxford University Press.
1224 25. Henrich J, McElreath R. The evolution of cultural evolution. Evolutionary Anthropology:
1225 Issues, News, and Reviews. 2003;12(3):123-35.
1226 26. Herrmann E, Call J, Hernández-Lloreda MV, Hare B, Tomasello M. Humans have evolved
1227 specialized skills of social cognition: the cultural intelligence hypothesis. Science.
1228 2007;317(5843):1360-6.
1229 27. Dean LG, Vale GL, Laland KN, Flynn E, Kendal RL. Human cumulative culture: a
1230 comparative perspective. Biological Reviews. 2014;89(2):284-301.
1231 28. Dean LG, Kendal RL, Schapiro SJ, Thierry B, Laland KN. Identification of the social and
1232 cognitive processes underlying human cumulative culture. Science. 2012;335(6072):1114-8.
1233 29. Lewis HM, Laland KN. Transmission fidelity is the key to the build-up of cumulative
1234 culture. Philosophical Transactions of the Royal Society B: Biological Sciences.
1235 2012;367(1599):2171-80.
1236 30. Lipschuetz M, Cohen SM, Ein-Mor E, Sapir H, Hochner-Celnikier D, Porat S, et al. A large
1237 head circumference is more strongly associated with unplanned cesarean or instrumental delivery
1238 and neonatal complications than high birthweight. American journal of obstetrics and gynecology.
1239 2015;213(6):833. e1-. e12.

1240 31. Dunbar RIM. The social brain hypothesis. Evolutionary Anthropology: Issues, News, and
1241 Reviews. 1998;6(5):178-90. doi: 10.1002/(SICI)1520-6505(1998)6:5<178::AID-
1242 EVAN5>3.0.CO;2-8.
1243 32. Dunbar RI. The social brain hypothesis and its implications for social evolution. Annals of
1244 human biology. 2009;36(5):562-72.
1245 33. Shultz S, Dunbar RI. The evolution of the social brain: anthropoid primates contrast with
1246 other vertebrates. Proceedings of the Royal Society of London B: Biological Sciences.
1247 2007;274(1624):2429-36.
1248 34. Fox KC, Muthukrishna M, Shultz S. The social and cultural roots of whale and dolphin
1249 brains. Nature ecology & evolution. 2017.
1250 35. McNally L, Brown SP, Jackson AL. Cooperation and the evolution of intelligence.
1251 Proceedings of the Royal Society of London B: Biological Sciences. 2012:rspb20120206.
1252 36. McNally L, Jackson AL. Cooperation creates selection for tactical deception. Proceedings
1253 of the Royal Society of London B: Biological Sciences. 2013;280(1762):20130699.
1254 37. Dávid-Barrett T, Dunbar R. Processing power limits social group size: computational
1255 evidence for the cognitive costs of sociality. Proceedings of the Royal Society B: Biological
1256 Sciences. 2013;280(1765):20131151.
1257 38. Gavrilets S. Collective action and the collaborative brain. J R Soc Interface. 2015;12(102).
1258 doi: 10.1098/rsif.2014.1067.
1259 39. González-Forero M, Gardner A. Inference of ecological and social drivers of human brain-
1260 size evolution. Nature. 2018;557(7706):554-7. doi: 10.1038/s41586-018-0127-x.
1261 40. Gavrilets S, Vose A. The dynamics of Machiavellian intelligence. Proceedings of the
1262 National Academy of Sciences. 2006;103(45):16823-8.

1263 41. Humphrey NK. The social function of intellect. In: Bateson PPG, Hinde RA, editors.
1264 Growing points in ethology. Cambridge, UK: Cambridge University Press; 1976. p. 303-17.
1265 42. Whiten A, Byrne RW. Taking Machiavellian intelligence apart. In: Byrne RW, Whiten A,
1266 editors. Machiavellian intelligence: social complexity and the evolution fo intellect in monkeys,
1267 apes and humans. Oxford, UK: Oxford University Press; 1988. p. 50-65.
1268 43. Whiten A, Byrne RW. The Machiavellian intelligence hypothesis. In: Byrne RW, Whiten
1269 A, editors. Machiavellian intelligence: social complexity and the evolution fo intellect in monkeys,
1270 apes and humans. Oxford, UK: Oxford University Press; 1988. p. 1-9.
1271 44. Jolly A. Lemur Social Behavior and Primate Intelligence. Science. 1966;153(3735):501-6.
1272 doi: 10.2307/1719878.
1273 45. Pradhan GR, Tennie C, van Schaik CP. Social organization and the evolution of cumulative
1274 technology in apes and hominins. Journal of human evolution. 2012;2012(1):1-11. doi:
1275 10.1016/j.jhevol.2012.04.008. PubMed PMID: 22658335.
1276 46. Reader SM, Laland KN. Social intelligence, innovation, and enhanced brain size in
1277 primates. Proceedings of the National Academy of Sciences. 2002;99(7):4436-41.
1278 47. Doebeli M, Hauert C, Killingback T. The evolutionary origin of cooperators and defectors.
1279 Science. 2004;306(5697):859-62.
1280 48. Henrich J. Demography and cultural evolution: how adaptive cultural processes can
1281 produce maladaptive losses: the Tasmanian case. American Antiquity. 2004;69(2):197-214.
1282 49. Henrich J, Boyd R, Derex M, Kline MA, Mesoudi A, Muthukrishna M, et al.
1283 Understanding cumulative cultural evolution. Proceedings of the National Academy of Sciences.
1284 2016. doi: 10.1073/pnas.1610005113.

1285 50. Vaesen K, Collard M, Cosgrove R, Roebroeks W. Population size does not explain past
1286 changes in cultural complexity. Proceedings of the National Academy of Sciences.
1287 2016;113(16):E2241-E7.
1288 51. Henrich J, Boyd M, Derex M, Kline MA, Mesoudi A, Muthukrishna M, et al.
1289 Understanding Cumulative Cultural Evolution. Proceedings of the National Academy of Sciences.
1290 in press.
1291 52. Barton RA. Neocortex size and behavioural ecology in primates. Proceedings of the Royal
1292 Society of London Series B: Biological Sciences. 1996;263(1367):173-7.
1293 53. Lefebvre L. Brains, innovations, tools and cultural transmission in birds, non-human
1294 primates, and fossil hominins. Frontiers in human neuroscience. 2013;7(245).
1295 54. Charvet CJ, Finlay BL. Embracing covariation in brain evolution: large brains, extended
1296 development, and flexible primate social systems. Progress in brain research. 2012;195:71.
1297 55. Isler K, van Schaik CP. The expensive brain: a framework for explaining evolutionary
1298 changes in brain size. Journal of Human Evolution. 2009;57(4):392-400.
1299 56. Joffe TH. Social pressures have selected for an extended juvenile period in primates.
1300 Journal of human evolution. 1997;32(6):593-605.
1301 57. Walker R, Burger O, Wagner J, Von Rueden CR. Evolution of brain size and juvenile
1302 periods in primates. Journal of Human Evolution. 2006;51(5):480-9.
1303 58. Clutton‐Brock T, Harvey PH. Primates, brains and ecology. Journal of zoology.
1304 1980;190(3):309-23.
1305 59. Harvey PH, Krebs JR. Comparing brains. Science. 1990;249:150-6.
1306 60. Rogers AR. Does biology constrain culture? American Anthropologist. 1988;90(4):819-
1307 31.

1308 61. Pianka ER. Evolutionary ecology: Eric R. Pianka; 2011.

1309 62. Hill K, Hurtado AM. Cooperative breeding in South American hunter–gatherers.
1310 Proceedings of the Royal Society B: Biological Sciences. 2009;276(1674):3863-70. PubMed
1311 PMID: 19692401; PubMed Central PMCID: PMCPMC2817285.

1312 63. Lansing JS, Watkins JC, Hallmark B, Cox MP, Karafet TM, Sudoyo H, et al. Male
1313 dominance rarely skews the frequency distribution of Y chromosome haplotypes in human
1314 populations. Proceedings of the National Academy of Sciences. 2008;105(33):11645-50. doi:
1315 10.1073/pnas.0710158105.

1316 64. Hooper PL, Ross CT, Mulder MB, al. e. Low male reproductive skew in humans compared
1317 to non-human mammals. Working Paper2016.

1318 65. Magnusson K. R Psychologist 2016. Available from: http://rpsychologist.com/.

1319 66. Fenner JN. Cross-cultural estimation of the human generation interval for use in genetics-
1320 based population divergence studies. American journal of physical anthropology.
1321 2005;128(2):415-23. doi: 10.1002/ajpa.20188.

1322 67. Kumar S, Filipski A, Swarna V, Walker A, Hedges SB. Placing confidence limits on the
1323 molecular age of the human–chimpanzee divergence. Proceedings of the National Academy of
1324 Sciences of the United States of America. 2005;102(52):18842-7.

1325 68. Chudek M, Muthukrishna M, Henrich J. Cultural Evolution. In: Buss DM, editor. The
1326 Handbook of Evolutionary Psychology. 2. 2nd ed. Hoboken, NJ: John Wiley and Sons; 2015.

1327 69. Rendell L, Boyd R, Cownden D, Enquist M, Eriksson K, Feldman MW, et al. Why copy
1328 others? Insights from the social learning strategies tournament. Science. 2010;328(5975):208-13.

1329 70. Kendal RL, Boogert NJ, Rendell L, Laland KN, Webster M, Jones PL. Social Learning
1330 Strategies: Bridge-Building between Fields. Trends in cognitive sciences. 2018.

1331 71. Evans KE, Harris S. Adolescence in male African elephants, Loxodonta africana, and the
1332 importance of sociality. Animal Behaviour. 2008;76(3):779-87.
1333 72. Olesiuk P, Bigg M, Ellis G. Life history and population dynamics of resident killer whales
1334 (Orcinus orca) in the coastal waters of British Columbia and Washington State. Report of the
1335 International Whaling Commission, Special. 1990;(12):209-43.
1336 73. Shultz S, Dunbar RI. Social bonds in birds are associated with brain size and contingent on
1337 the correlated evolution of life‐history and increased parental investment. Biological Journal of
1338 the Linnean Society. 2010;100(1):111-23.
1339 74. Pérez‐Barbería FJ, Shultz S, Dunbar RI. Evidence for coevolution of sociality and relative
1340 brain size in three orders of mammals. Evolution. 2007;61(12):2811-21.
1341 75. Hawn AT, Radford AN, du Plessis MA. Delayed breeding affects lifetime reproductive
1342 success differently in male and female green woodhoopoes. Current biology. 2007;17(10):844-9.
1343 76. Demenocal PB. Plio-pleistocene African climate. Science. 1995:53-9.
1344 77. Kline MA, Boyd R. Population size predicts technological complexity in Oceania.
1345 Proceedings of the Royal Society B: Biological Sciences. 2010;277(1693):2559-64.
1346 78. Powell A, Shennan S, Thomas MG. Late Pleistocene demography and the appearance of
1347 modern human behavior. Science. 2009;324(5932):1298-301.
1348 79. Kobayashi Y, Aoki K. Innovativeness, population size and cumulative cultural evolution.
1349 Theoretical Population Biology. 2012;82(1):38-47.
1350 80. Enquist M, Strimling P, Eriksson K, Laland K, Sjostrand J. One cultural parent makes no
1351 culture. Animal Behaviour. 2010;79(6):1353-62.
1352 81. Henrich J, Tennie C. Cultural Evolution in Chimpanzees and Humans. forthcoming. In:
1353 Chimpanzees and Human Evolution [Internet]. Cambridge: Harvard University Press.

1354 82. Morgan TJH, Uomini NT, Rendell LE, Chouinard-Thuly L, Street SE, Lewis HM, et al.
1355 Experimental evidence for the co-evolution of hominin tool-making teaching and language. Nature
1356 Communications. 2015;6.
1357 83. Fiddes IT, Lodewijk GA, Mooring M, Bosworth CM, Ewing AD, Mantalas GL, et al.
1358 Human-specific NOTCH2NL genes affect Notch signaling and cortical neurogenesis. Cell.
1359 2018;173(6):1356-69. e22.
1360 84. Suzuki IK, Gacquer D, Van Heurck R, Kumar D, Wojno M, Bilheu A, et al. Human-
1361 specific NOTCH2NL genes expand cortical neurogenesis through Delta/Notch regulation. Cell.
1362 2018;173(6):1370-84. e16.
1363 85. MacLean EL, Hare B, Nunn CL, Addessi E, Amici F, Anderson RC, et al. The evolution
1364 of self-control. Proceedings of the National Academy of Sciences of the United States of America.
1365 2014;111(20):E2140-E8. PubMed PMID: PMC4034204.
1366 86. Benson-Amram S, Dantzer B, Stricker G, Swanson EM, Holekamp KE. Brain size predicts
1367 problem-solving ability in mammalian carnivores. Proceedings of the National Academy of
1368 Sciences. 2016;113(9):2532-7.
1369 87. Holekamp KE, Benson-Amram S. The evolution of intelligence in mammalian carnivores.
1370 Interface Focus. 2017;7(3):20160108.
1371 88. Street SE, Navarrete AF, Reader SM, Laland KN. Coevolution of cultural intelligence,
1372 extended life history, sociality, and brain size in primates. Proceedings of the National Academy
1373 of Sciences. 2017;114(30):7908-14.
1374 89. Moll H, Tomasello M. Cooperation and human cognition: the Vygotskian intelligence
1375 hypothesis2007 2007-04-29 00:00:00. 639-48 p.

1376 90. Ruff CB, Trinkaus E, Holliday TW. Body mass and encephalization in Pleistocene Homo.
1377 Nature. 1997;(387):173-6.
1378

[i] Curves generated using 65. Magnusson K. R Psychologist 2016. Available from: http://rpsychologist.com/. (rpsychologist.com)

Supporting Information Legends

CBH_Supplementary.pdf. CBH Analytical Model. Full details for the Cultural Brain Hypothesis adaptive dynamics model.

CBH_Code.tar.gz. CBH Simulation Model. C++ code for the Cultural Brain Hypothesis simulation model.

Supplementary Material
to
The Cultural Brain Hypothesis: How culture drives brain expansion, sociality, and life history

Michael Muthukrishna, Michael Doebeli, Maciej Chudek & Joseph Henrich

Analytical Model

To explore the evolutionary adaptive dynamics of the CBH, we begin with individuals i represented by three continuous variables: brain size b_i, adaptive knowledge a_i, and reliance on social learning (over asocial learning; e.g. time spent), s_i. We will initially ignore the evolution of oblique learning, learning biases, and population structure, and assume that individuals using social learning use oblique learning and learning biases to hone in on the target individual with the most adaptive knowledge. We will relax this assumption in our simulation and allow oblique learning, learning biases, life history, and population structure to endogenously evolve. Table 1 is a handy key for the variables in our analytic model.

Table 1. Variables for analytic model.

Variable	Description	Values
b_i	Brain size of individual i	$[0, \infty]$
a_i	Adaptive knowledge of individual i	$[0, b_i]$
s_i	Reliance on social learning (over asocial learning) of individual i	$[0,1]$
τ	Transmission fidelity	$[0,1]$
ζ	Asocial learning efficacy	$[0,1]$
r	Per capita birth rate	≥ 0
d	Per capita death rate	≥ 0
ϕ	Scaling coefficient used to scale relationship between adaptive knowledge and r	≥ 0
β	Scaling constant relating brain size to death rate (similar to κ in simulation)	≥ 0
λ	Death rate mitigation (e.g. richness of ecology) used to scale effect of a_i in reducing d	≥ 0
n	Increase in population size	$[-\infty, \infty]$

Individual i has two routes to acquire adaptive knowledge a_i: (1) through asocial (individual) learning as a function of their own brain size b_i and (2) through social learning as a function of the target model from whom they learn (a_m). The proportion of time or propensity to use social over asocial learning is given by s_i. Thus adaptive knowledge is given by:

$$a_i = s_i \cdot \tau \cdot a_m + (1 - s_i) \cdot \zeta \cdot b_i \quad (1)$$

Where τ is transmission fidelity (how well an individual can learn from a model), ζ is asocial learning efficacy (how effectively an individual can use their brain to figure things out), and a_m is the adaptive knowledge possessed by the individual in the parent generation from whom they are learning (e.g. model with maximal a or model with average a, etc).

The parameters τ and ζ in Equation 1 are abstractions of more complicated details covered in other work. By outsourcing the evolution of these features to other models, we can focus on the core of the CBH argument; i.e. how learning, brain size, knowledge, sociality, and life history are interconnected. Examples of this earlier work includes, Lewis and Laland [1] model of the relationship between transmission fidelity and the rate of trait loss, showing that sufficiently high transmission fidelity is necessary for cumulative culture, even more so than novel invention, incremental improvement, and recombination. Relatedly, building on work by Henrich [2], Mesoudi [3] models how increases in cumulative culture (driven by, for example, sociality) are more difficult for each generation to acquire. Thus, selection favors mechanisms to increase transmission fidelity. Muthukrishna and Henrich [4] discuss the many mechanisms to increase transmission fidelity as adaptive knowledge accumulates. Mechanisms such as explicit teaching may not be required in a small-scale society, but in a large-scale society, not only is explicit teaching required, but also formal institutionalized schooling from a variety of teachers. Thus, τ could include individuals' cognitive abilities [itself increased by culture; see 4], but also greater social tolerance, more interactions or opportunities for interaction, and some passive or active teaching by models, and so on [for more examples, see 4, 5, 6]. Transmission fidelity could be broken down into constraints and endogenous state variables for genetic, cultural, and social factors, as well as interactions between these (e.g. genes for sociality), but for the purposes of expressing our argument, here we capture all this with τ.

Similarly, our model relies on the idea that "bigger" brains will be better at solving novel problems, and figuring stuff out [7, 8]. As Deaner, et al. [7] analyses reveal, at least in primates, the best predictor of cognitive ability is overall brain size. But, as with transmission fidelity, many factors will influence individuals' ability to use their brains, such as constraints on time (for trial and error learning) or energy. These constraints are captured by ζ.

We take an evolutionary adaptive dynamics approach to find the evolutionary stable strategies (ESS) in our model. This approach involves assuming a monomorphic population and then looking at the "invasibility" of the population to a mutant (in variables of interest) with slightly different values. Appropriate to the dynamics we are interested in, this analytic method assumes mutations are small (i.e. we are not exploring competition between two vastly different groups).

Social Learning. To determine the average adaptive knowledge in a population that is monomorphic for resident genotype (s, b), we'll initially assume that genotype is fixed over the course of learning. We'll assume that the learning process leads to a distribution of adaptive knowledge values in the population and that individuals using social learning select a model using payoff-biased learning, choosing to learn from the model with the maximal possible value of adaptive knowledge a_{max} (i.e., they learn from the rare individual who has attained the maximal value). In the simulation, we will relax this assumption and allow oblique learning (learning from

non-genetic parents) and learning bias to evolve. Assuming individuals do learn from the best model when social learning, the mean adaptive knowledge in the population is given by:

$$\bar{a}(s,b) = s \cdot \tau \cdot a_{max} + (1-s) \cdot \zeta \cdot b \quad (2)$$

We further assume that the maximal adaptive knowledge is constrained by the brain size of the learner, such that $a_{max} = \nu b$, where $\nu > 0$ is some scaling parameter. As we shall see, the insights of the model are independent of the specific value of the ν scaling. Thus, Equation 2 becomes:

$$\bar{a}(s,b) = s \cdot \tau \cdot \nu \cdot b + (1-s) \cdot \zeta \cdot b \quad (2a)$$

We can now easily understand the adaptive dynamics of the social learning trait (s) assuming more adaptive knowledge has a higher payoff. For a given brain size (b), we simply compare $\tau\nu b$ and ζb: if $\tau\nu b > \zeta b$, then it pays to increase s as much as possible to maximize adaptive knowledge (i.e. $s \to 1$); conversely, if $\tau\nu b < \zeta b$, then it pays to decrease s as much as possible to maximize adaptive knowledge (i.e. $s \to 0$). This will be true as long as individuals have access to a range of models and are learning from the model with the greatest adaptive knowledge. Given these conditions, the key to reliance on social learning is the ability to learn with high fidelity and the key to reliance on asocial learning is the ability to efficiently use one's brain to learn by oneself. Further, if there is some limitation on accessing the model with the maximal adaptive knowledge, such as ineffective payoff biased learning making it difficult to identify who has the most adaptive knowledge or too small or disconnected a population for at least one individual to consistently reach this maximal value every generation, then the evolution of social learning is also going to depend on the maximal adaptive knowledge learners have access to. We explore these dynamics in the simulation model.

Brain Size. To determine the adaptive dynamics of brain size, we need an ecological model for monomorphic populations (i.e. for populations that consist of a single resident type (s, b)). To do this, we need to specify how the various traits affect the birth and death rates in the model. We use a logistic ecological model:

$$\frac{dN}{dt} = (r-d) \cdot N \quad (3)$$

Here N is population density, r is the per capita birth rate of the resident and d is the per capita death rate of the resident. Next, we specify the per capita birth rate (r) and death rate (d). We assume the birth rate r decreases with population size (density dependence influencing carrying capacity), but that that decrease is slower with increased adaptive knowledge (e.g. allowing you to support more offspring or outcompete competitors in access to mating opportunities). The birth rate (r) is given by:

$$\begin{aligned} r &= \rho\left(1 - \frac{N}{k_0 + k_1\bar{a}}\right) \\ &= \rho\left(1 - \frac{N}{k_0 + k_1(s \cdot \tau \cdot \nu \cdot b + (1-s) \cdot \zeta \cdot b)}\right) \end{aligned} \quad (4)$$

Where ρ is the maximal birth rate and that dependence leads to a linear decrease in the birth rate given by the second half of Equation 4. This linear decrease is assumed to be influenced by the mean adaptive knowledge $(\bar{a})$, such that more adaptive knowledge leads to a larger denominator, slowing the decrease with density dependence (allowing for a higher effective carrying capacity). k_0 and k_1 are positive parameters, which we set to 1, without loss of generality, in the following analyses.

We assume that a larger brain is more costly than a smaller brain in terms of death rate (e.g. higher calorie requirements), but that more adaptive knowledge lowers the death rate (e.g. finding food or evading predators). The death rate (d) is given by:

$$d = \beta \cdot b^n \cdot e^{-\lambda a/b} \quad (5)$$

This function assumes that the cost of brains scales up in a polynomial fashion (e.g. $n = 2$), but that the reduction in the death rate through adaptive knowledge is an exponential decay, where adaptive knowledge is bounded by brain size (i.e. $a \leq \nu b$). Here β scales the maximum brain size and λ scales the death rate reducing payoff to adaptive knowledge. The degree to which adaptive knowledge can offset brain size is determined by λ and the ratio of adaptive knowledge to brain size (adaptive knowledge is constrained by brain size regardless of learning mechanism and as brains grow, more knowledge is required to provide an equivalent offset). The λ parameter allows us to adjust the extent to which adaptive knowledge can offset the costs of brain size, where $\lambda = 0$ indicates no offset and increasing λ increases the probability of survival for a given adaptive knowledge and brain size. The λ parameter can be interpreted as how much adaptive knowledge one requires to unlock the fitness-enhancing advantages. For example, in a calorie-rich environment where only a little skill or knowledge is required to access calories (e.g. simply remembering food locations), λ would be high – a little bit of knowledge gives a large return. Conversely, in a calorie-poor environment where a lot of skills or knowledge are required to access fewer calories (e.g. food needs significant preparation before safe consumption), λ would be low. Note that this is a mechanical relationship between adaptive knowledge and probability of survival. Calorie availability is a potentially useful metaphor to think about the model in concrete term, but its is not the only interpretation of λ, which could be influenced any number of factors, including knowledge required to evade predators or avoid environmental hazards. We are also not directly saying anything about selection on cognition, brain size, or information use [9]. For example, high λ might allow for larger brains due to greater food availability for given food finding knowledge, but equally, when λ is low, there may be selection pressure for larger brains with more knowledge needed to acquire the more difficult to access food.

In the analytic model, the decrease to the death rate through adaptive knowledge becomes a constant since adaptive knowledge is a function of brain size (and parameters affecting learning efficiency), but although this will not affect the dynamics of the model, it will affect the final brain sizes. We fully explore this in the simulation.

Given a resident (s, b), the equilibrium population size of the resident is determined by the solution to Equation 3:

$$r(N^*) = d(b) \quad (6)$$

Since we know that $s \to 0$ when $\tau\nu < \zeta$ and $s \to 1$ when $\tau\nu > \zeta$, we can consider these two cases, asocial learners and social learners, separately and then compare the outcomes of these two regimes.

Asocial learners ($s = 0$)

To determine the adaptive dynamics of brain size, consider a mutant (designated by subscript "m") with brain b_m. This mutant's adaptive knowledge based on Equation 1 will be $a_m = \zeta b_m$, since $s = 0$. Using the same ecological assumptions as before for a mutant type b_m, and assuming the mutant is rare and growing (initially) in a resident population that is at its ecological equilibrium N^*, the per capita growth rate of the mutant, its invasion fitness, is:

$$
\begin{aligned}
f(b, b_{\mathrm{m}}) &= r_m(N^*) - d(b_m) \\
&= \rho\left(1 - \frac{N^*}{1 + a_m}\right) - \beta \cdot b_m^n \cdot e^{-\lambda^{a_m}/_{b_m}} \\
&= \rho\left(1 - \frac{N^*}{1 + \zeta b_m}\right) - \beta \cdot b_m^n \cdot e^{-\lambda^{\zeta b_m}/_{b_m}} \\
&= \rho\left(1 - \frac{N^*}{1 + \zeta b_m}\right) - \beta \cdot b_m^n \cdot e^{-\lambda\zeta}
\end{aligned} \tag{7}
$$

To examine the adaptive dynamics of brain size, we need to calculate the selection gradient by taking the derivative of the invasion fitness f with respect to the mutant trait b_m and evaluate this derivative at the resident value b. To calculate if these equilibria are stable, we will calculate the second derivative. If the second derivative is negative, then the value is a convergent stable ESS. For those unfamiliar with this approach, it may be helpful to use a physical analog—distance, speed, and acceleration (or more accurately, displacement, velocity, and acceleration). The derivative of distance over time (metres) is speed (metres per second). The second derivative (derivative of speed) is acceleration (metres per second per second). The adaptive dynamics approach is the equivalent of looking at when an object is stationary (i.e. speed—derivative of distance—is 0) and confirming that these "equilibria" stationary points are convergent by confirming that objects decelerate around these points (i.e. acceleration—second derivative—is negative). If the second derivative were positive, objects would increase speed and move away from this stationary point, or in the present case, there would be positive selection for mutants away from this equilibrium. Let us calculate the selection gradient for brain size:

$$
\frac{db}{dt} = \frac{\delta f}{\delta b_m}|_{b_m = b} = \frac{\rho\zeta N^*}{(1 + \zeta b)^2} - n\beta e^{-\lambda\zeta} b^{n-1} \tag{8}
$$

From Equation 8 we can see that if $n > 1$, $db/dt < 0$ for large b and $db/dt > 0$ for small b, which suggests that there is some intermediate ESS value for brain size (b^*). It is straightforward to check that the second derivative of the invasion fitness function (Equation 8) with respect to the mutant trait and evaluated at the resident trait is always negative and therefore the singular strategy b^* is a CSS (i.e., a convergent stable ESS). This equilibrium brain value (i.e. when $db/dt = 0$) is difficult to solve for a generic polynomial n. To calculate a solution, we can select a reasonable polynomial (e.g. $n = 2$, which we use in the simulation) and solve for $db/dt = 0$. As long as brain size is positive, the relationship between brain size and the death rate will be superlinear and monotonous; our qualitative results should be robust to the specific polynomial used. Here is the equilibrium brain size for $n = 2$:

$$
b^* = \frac{-\beta + \sqrt{\beta^2 + 3\rho\zeta^2\beta e^{\lambda\zeta}}}{3\zeta\beta} \tag{9}
$$

We need to compare the equilibrium brain size among asocial learners expressed in Equation 9 with the equilibrium brain size among social learners, so let's now calculate the dynamics for social learners.

Social learners ($s = 1$)

To determine the adaptive dynamics of brain size, consider a mutant (designated by subscript "m") with brain b_m. This mutant's adaptive knowledge based on Equation 1 will be $a_m = \tau v b_m$, since

$s = 1$. Using the same ecological assumptions as before for a mutant type b_m, and assuming the mutant is rare and growing (initially) in a resident population that is at its ecological equilibrium N^*, the per capita growth rate of the mutant, its invasion fitness, is:

$$\begin{aligned} f(b_{\text{res}}, b_{\text{m}}) &= r_m(N^*) - d(b_m) \\ &= \rho\left(1 - \frac{N^*}{1 + a_m}\right) - \beta \cdot b_m^n \cdot e^{-\lambda^{a_m}/_{b_m}} \\ &= \rho\left(1 - \frac{N^*}{1 + \tau v b_m}\right) - \beta \cdot b_m^n \cdot e^{-\lambda^{\tau v b_m}/_{b_m}} \\ &= \rho\left(1 - \frac{N^*}{1 + \tau v b_m}\right) - \beta \cdot b_m^n \cdot e^{-\lambda \tau v} \end{aligned} \quad (10)$$

As before, to examine the adaptive dynamics of brain size, we need to calculate the selection gradient by taking the derivative of the invasion fitness f with respect to the mutant trait b_m and evaluate this derivative at the resident value b_{res}. To calculate if these equilibria are stable, we will calculate the second derivative. If the second derivative is negative, then the value is a convergent stable ESS. Let us calculate the selection gradient for the brain size of social learners:

$$\frac{db}{dt} = \frac{\delta f}{\delta b_m}|_{b_m = b_{res}} = \frac{\rho \tau v N^*}{(1 + \tau v b)^2} - n\beta e^{-\lambda \tau v} b^{n-1} \quad (11)$$

As with asocial learners, from Equation 11 we can see that if $n > 1$, $db/dt < 0$ for large b and $db/dt > 0$ for small b, which suggests that there is some intermediate ESS value for brain size (b^*). As before, it is straightforward to check that the second derivative of the invasion fitness function (Equation 11) with respect to the mutant trait and evaluated at the resident trait is always negative and therefore the singular strategy b^* is a CSS (i.e., a convergent stable ESS). We can set $n = 2$ and calculate this equilibrium brain value (i.e. when $db/dt = 0$):

$$b^* = \frac{-\beta + \sqrt{\beta^2 + 3\rho\tau^2 v^2 \beta e^{\lambda\zeta}}}{3\tau v \beta} \quad (12)$$

Equation 12 is functionally similar to Equation 9, but the equilibrium brain size for asocial and social learners will be different. Moreover, since to enter the realm of social learning, $\tau v b > \zeta b$, social learners, *ceteris paribus*, will have larger equilibrium brain sizes than asocial learners. Note this prediction – that social learners will have larger brain sizes than asocial learners – is an outcome of the model, not an assumption. Moreover, transmission fidelity, asocial learning efficacy, and the payoff for adaptive knowledge (e.g. richness of the environment) are all going to affect the equilibrium brain size.

References

1. Lewis HM, Laland KN. Transmission fidelity is the key to the build-up of cumulative culture. Philosophical Transactions of the Royal Society B: Biological Sciences. 2012;367(1599):2171-80.
2. Henrich J. Demography and cultural evolution: how adaptive cultural processes can produce maladaptive losses: the Tasmanian case. American Antiquity. 2004;69(2):197-214.
3. Mesoudi A. Variable cultural acquisition costs constrain cumulative cultural evolution. PloS one. 2011;6(3):e18239.

4. Muthukrishna M, Henrich J. Innovation in the collective brain. Philosophical Transactions of the Royal Society of London B: Biological Sciences. 2016;371(1690). doi: 10.1098/rstb.2015.0192.
5. Dean LG, Kendal RL, Schapiro SJ, Thierry B, Laland KN. Identification of the social and cognitive processes underlying human cumulative culture. Science. 2012;335(6072):1114-8.
6. Whiten A, Erdal D. The human socio-cognitive niche and its evolutionary origins. Philosophical Transactions of the Royal Society of London B: Biological Sciences. 2012;367(1599):2119-29.
7. Deaner RO, Isler K, Burkart J, van Schaik C. Overall brain size, and not encephalization quotient, best predicts cognitive ability across non-human primates. Brain, Behavior and Evolution. 2007;70(2):115-24.
8. Sol D, Bacher S, Reader SM, Lefebvre L. Brain size predicts the success of mammal species introduced into novel environments. The American Naturalist. 2008;172(S1):S63-S71.
9. Dall SR, Giraldeau L-A, Olsson O, McNamara JM, Stephens DW. Information and its use by animals in evolutionary ecology. Trends in ecology & evolution. 2005;20(4):187-93.
10. Henrich J, Boyd R, Derex M, Kline MA, Mesoudi A, Muthukrishna M, et al. Understanding cumulative cultural evolution. Proceedings of the National Academy of Sciences. 2016. doi: 10.1073/pnas.1610005113.