Supplemental Material

Contents:

Movies

Four movies have been made available on YouTube (links provided below) for further visualization of the underlying model dynamics on the level of populations and generations. The movies complement Figure 3 in the main Article.

As the movies begin, the parameter scenario is first presented before the model display appears and the simulation starts running.

At the top of the display we find the individuals in the population represented as small circles in "α-space": the horizontal location of an individual i is determined by α_i. The display is updated with each generation, adapting the displayed interval of α, such that all individual values of α_i fit into view.

The individuals that were picked as role models for the coming generation are displayed in red, allowing the viewer to discern, for example, the effect of applying a more strict selection (higher values of s,) which will cause the red circles to gravitate toward the right (higher values of α.)

Bottom-left, a number of parameters and measures are displayed and updated as the simulation runs. We see the main parameters q, s and N, along with t, which denotes the current generation time step. We also see the average error (loss in imitation, λ; see Equation 4 in the main Article) and the average value of α_i among role models and among the whole population.

Bottom-right we see two plots that are filled in as the simulation proceeds: The top plot displays the evolution of average α in the population. The bottom plot displays the evolution of the parameter that is varied during the run.

Movie 1: Increasing N

This movie displays the evolution of a population as group size N is increased by a factor of 4 every 100 generations, beginning at a small population size of N = 4. We see how increasing N initially has a strong impact on average α, but that the effect abates the large the population gets. See text referring to Figure 3b in the main Article and Figure 13 in Appendix D in Supplemental Material.

URL: https://www.youtube.com/watch?v=NfHSaBCLawc

Movie 2: Increasing s

This movie displays the evolution of a population as selection strength s is increased by .2 every 100 generations, beginning at s = .1 (very weak selection for role models with large values of α). We see how increasing s has a strong and consistent impact on average α. See text referring to Figure 3c in the main Article and Figure 14 in Appendix D in Supplemental Material.

URL: https://www.youtube.com/watch?v=JsyEmgcrefg

Movie 3: Increasing q

This movie displays the evolution of a population as the imitation fidelity norm q is increased by .075 every 100 generations, beginning at q = .675. Increasing q has a strong and accelerating impact on average α_i, and the last bump (from q = .9 to q = .975) produces a jump in average α_i that is so large that the population hardly has time to settle down in the remaining 100 generations. See text referring to Figure 3d in the main Article and Figure 12 in Appendix D in Supplemental Material.

URL: https://www.youtube.com/watch?v=J0UdpmOkIo0

Movie 4: Reducing N

We here begin at an even larger population (N = 5,120) than Movie 1 ended at, reducing the population by a factor 4 every 100 generations. The idea here is to investigate a scenario that better represents the split-up of a large meta-population, in a way similar to what Henrich (2004) proposed for the Tasmanian case. We find that no conspicuous reduction of average α_i results as the population keeps getting quartered in this manner. There is a downward trend, but we must keep in mind that the rate of reduction is dramatic: at N = 20, average $\alpha_i \approx 17$, while at N = 5,120, average $\alpha_i \approx 21$. See also Figure 14d in Appendix D in Supplemental Material.

URL: https://www.youtube.com/watch?v=gHmK9TRo8aU

Appendix A: The Treadmill Model

Using the Price Equation (Price 1970), Henrich (2004) derives the mean change in skillfulness in one generation as

$$\Delta\bar{z} = -\alpha + \beta(\epsilon + Ln\ N) \qquad (7)$$

where α is the complexity of the skill, N is the number of "interacting social learners" and ϵ is the Euler-Gamma constant; see Equation 2 in Henrich (2004) for derivation. If $\Delta\bar{z} < 0$ this means that the skillfulness with which the skill is performed will deteriorate, leading eventually to the elimination of the skill.

The Treadmill Model yields two critical threshold values, one in α and one in N (see Henrich, 2004, for derivations):

$$\alpha_T^* = \beta\epsilon + \beta Ln\ N \qquad (8)$$

and

$$N_T^* > e^{\frac{\alpha}{\beta}-\epsilon}. \qquad (9)$$

Equation 8 gives us the maximum sustainable complexity of skills in a population of size N, and Equation 9 gives us the minimum population size needed to sustain a skill of complexity α. In Figure 8 we use Equation 7 to visualize these two critical values.

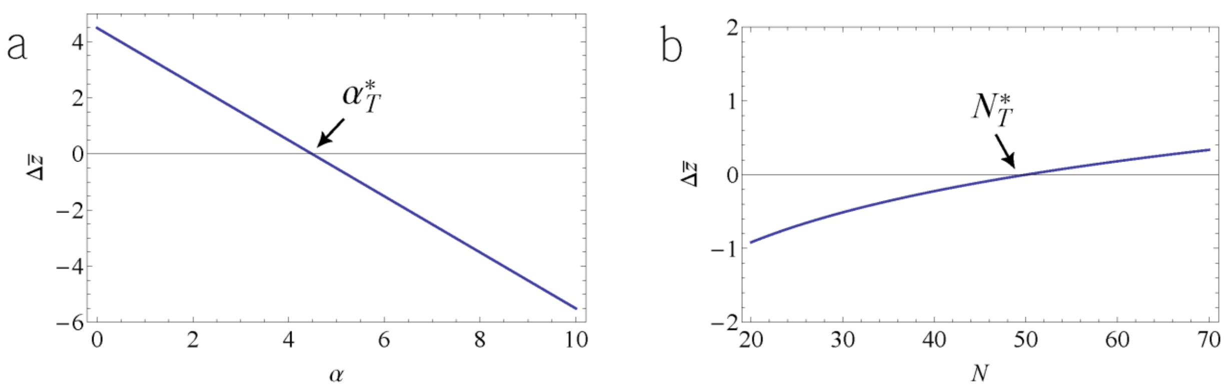

Figure 8: (a) High values of α will cause $\Delta\bar{z}$ to drop below 0, leading to deterioration and eventual loss of the skill. We plot Equation 7 (in Appendix A in Supplementary Materials) using β = 1 and N = 50, yielding $\alpha_T^* \approx 4.5$; see Equation 8 (in Appendix A in Supplementary Materials). (b) Low values of N have the same effect. Here we plot Equation 7 (in Appendix A in Supplementary Materials) using β = 1 and α =4.5, yielding $N_T^* \approx 50$; see Equation 9.

Appendix B: The Glass Ceiling Model

The central results of the Glass Ceiling Model can also be expressed as two critical values, one in α and one in q (see Fig. 9):

$$\alpha_G^* = -\frac{1}{Ln\, q}, \tag{10}$$

$$q_G^* = e^{-\frac{1}{\alpha}}; \tag{11}$$

see Andersson (2011) and Andersson (2013) for derivations of these equations, including the relation between the Glass Ceiling Model and the original Quasispecies Model.

The point of critical complexity emerges clearly in simulations since it becomes a stable equilibrium of the evolutionary system: selection will constantly drive α upward while the Glass Ceiling effect will cull skills with $\alpha > \alpha_G^*$ from above; see Andersson (2011:Figure 1) and Andersson (2013:Figure 2).

Population size has a very low impact on equilibrium α achieved in simulations of the Glass Ceiling (see Andersson 2013:Figure 8), but very low population counts are found by Andersson (2013) to cause instability (proneness to crashes), indicating the need for stabilizing cultural and cognitive pedagogical adaptations in groups of sizes ranging in the tens.

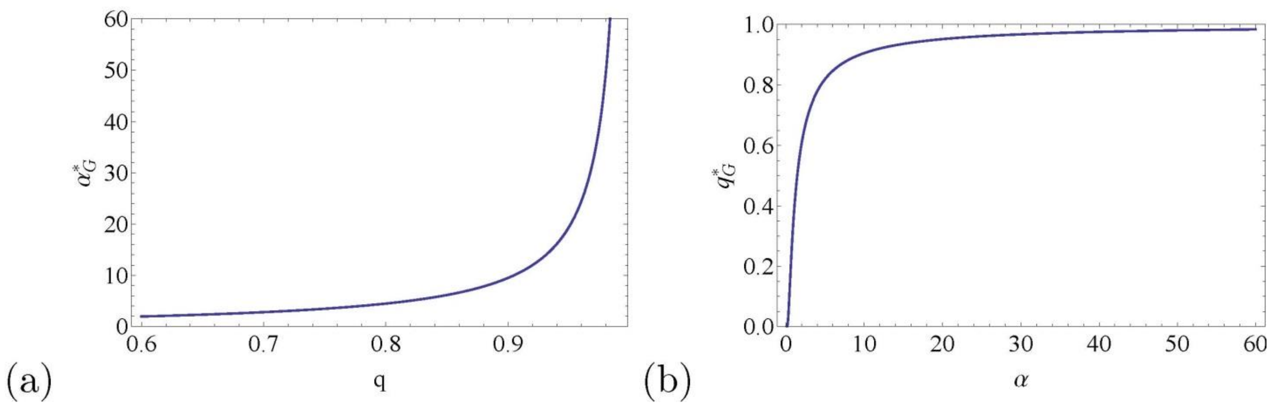

Figure 9: (a) Increasing q has a strong impact on the amount of complexity that can be maintained by imitation in the population; see α_G^* (Equation 10 in Appendix B in Supplementary Materials). (b) The level of imitation fidelity needed to maintain skills of a certain level of complexity α illustrates the level of the "Glass Ceiling"; q_G^* (Equation 11 in Appendix B in Supplementary Materials)

Appendix C. The Synthetic Model

We use a normal distribution of imitation outcomes rather than a Gumbel distribution, as used by Henrich (2004). The Gumbel distribution was used by Henrich for the specific purpose of building the strongest possible skill bias into the selection process, and its use also for imitation outcomes was an empirically unmotivated sacrifice for maintaining analytical tractability. The Treadmill Model has, as Henrich argued, since been further verified to be reasonably robust to this choice (see Vaesen 2012).

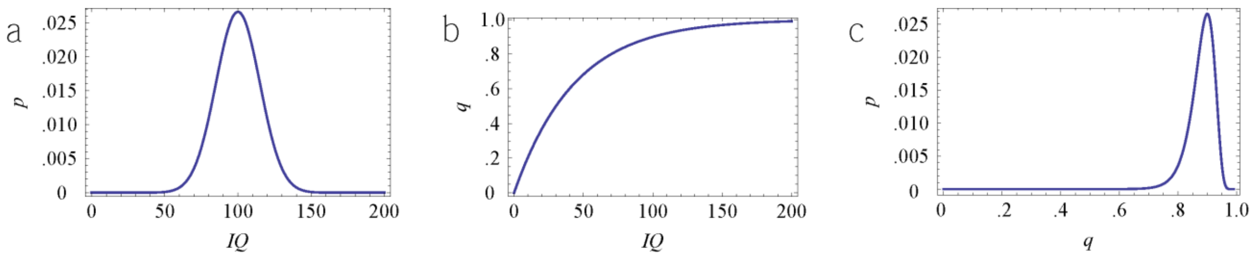

Figure 10: The Probability Density Function (PDF) of imitation fidelity q (c) is derived by combining the PDF of IQ values (a) with the mapping between q and values of IQ (b).

Lacking suitable empirical data, a model for how imitation fidelity q_i is distributed is derived by starting from the distribution of general intelligence in the population. General intelligence, as measured by IQ tests, is known to have a normal distribution, with parameters that are usually calibrated to be close to $\mu = 100$ and $\sigma = 15$ (Flynn, 1987). Assuming that imitation fidelity q would be correlated with general intelligence (ceteris paribus), we see it reasonable to assume that $q = 0$ for $IQ = 0$ and that $q \to 1$ as $IQ \to \infty$. The parameter q that we use to vary the general level of imitation fidelity in the population is then the "fidelity norm," i.e. transmission fidelity at IQ = 100.

Using IQ values sampled from this distribution, and the fidelity norm q, we use

$$q_i = 1 - e^{-\frac{1}{\mu} Ln\,(1-q) IQ_i}, \tag{12}$$

which has the sought-for properties; see Figure 10.

As in both models that go into the synthesis, role model selection is skill biased, but by contrast with Henrich (2004) we employ a parameter s by which we may set the strength of the skill bias between $s \in [0,1)$, where $s = 0$ corresponds to pure drift and where the probability that the most highly skilled role model is selected asymptotically approaches 1 as $s \to 1$.

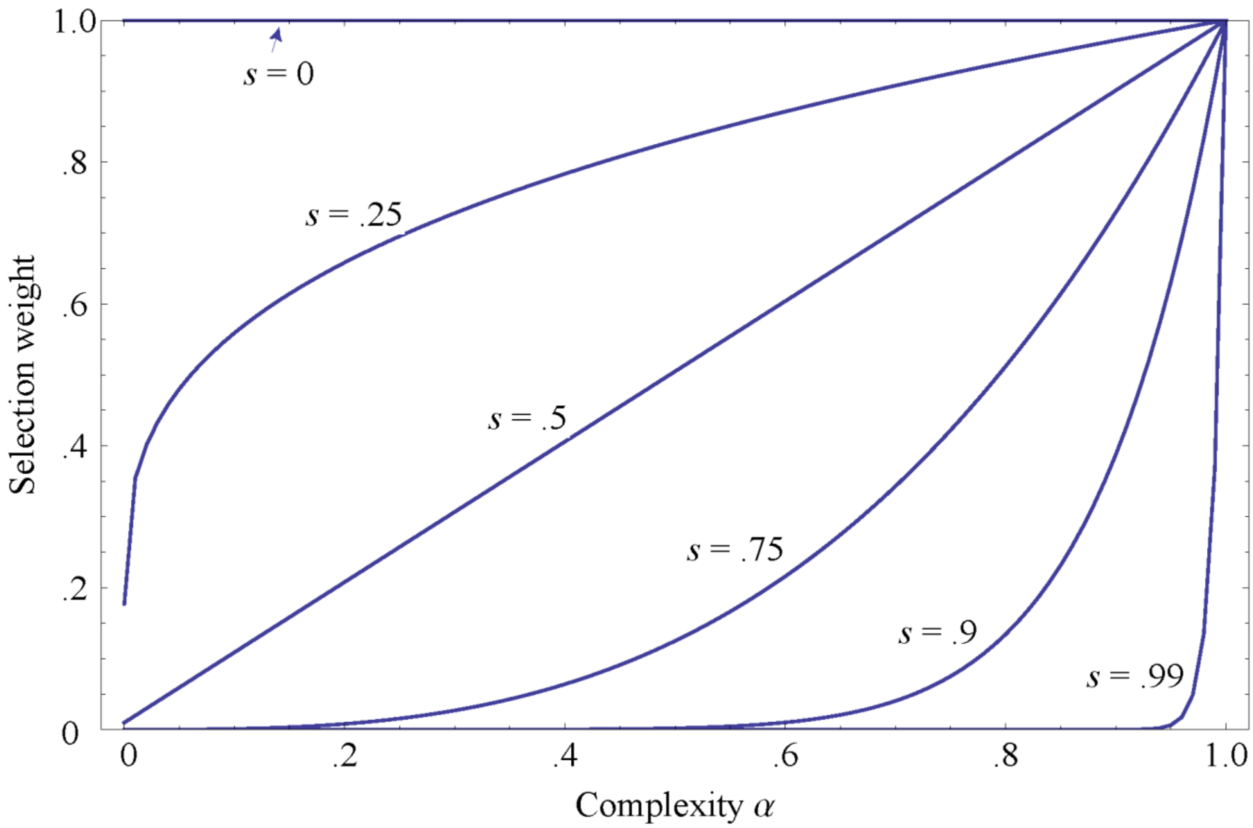

Figure 11: The effect of skill bias strength s on selection weight (y-axis) as a function of complexity α (x-axis) in the unit interval (assuming $\alpha_{max} = 1$, $\alpha_{min} = 0$; see Equation 13 in Appendix C in Supplementary Materials). At $s = 0$ selection weight is independent of skill complexity, corresponding to drift. Increasing s yields an accentuation of how skill complexity results in selection weight.

Updates are performed by letting a new generation of naive learners select role models to imitate among an older generation of encultured individuals; i.e. transmission is oblique (across generations but with no regard to genealogical relation) and synchronous. Role models are selected randomly using the relative magnitudes of the fitness value of the potential role models,

$$f_i = \frac{(\alpha_i - \alpha_{min})^{\frac{1}{1-s}-1} + 1}{(\alpha_{imax} - \alpha_{min})^{\frac{1}{1-s}-1} + 1}, \tag{13}$$

as the probability of being selected. Here, α_{max} represents the variant with the highest value of α_i in the population, and α_{min} the variant with the lowest α_i; see Figure 11.

The imitation event generates a complexity α_i from a normal distribution, see Equation 5 and Equation 6. As can be seen in Figure 2, as α increases under positive selection, the distance between the mean of the distribution and the α value of selected role models increases. This penalty for large values of α can be counteracted by larger values of q in the imitators.

See also the ODD specification provided below.

Appendix D. Response of the Synthetic Model to Historical Parameter Changes

We here explore responses to historical change events in all three parameters under investigation.

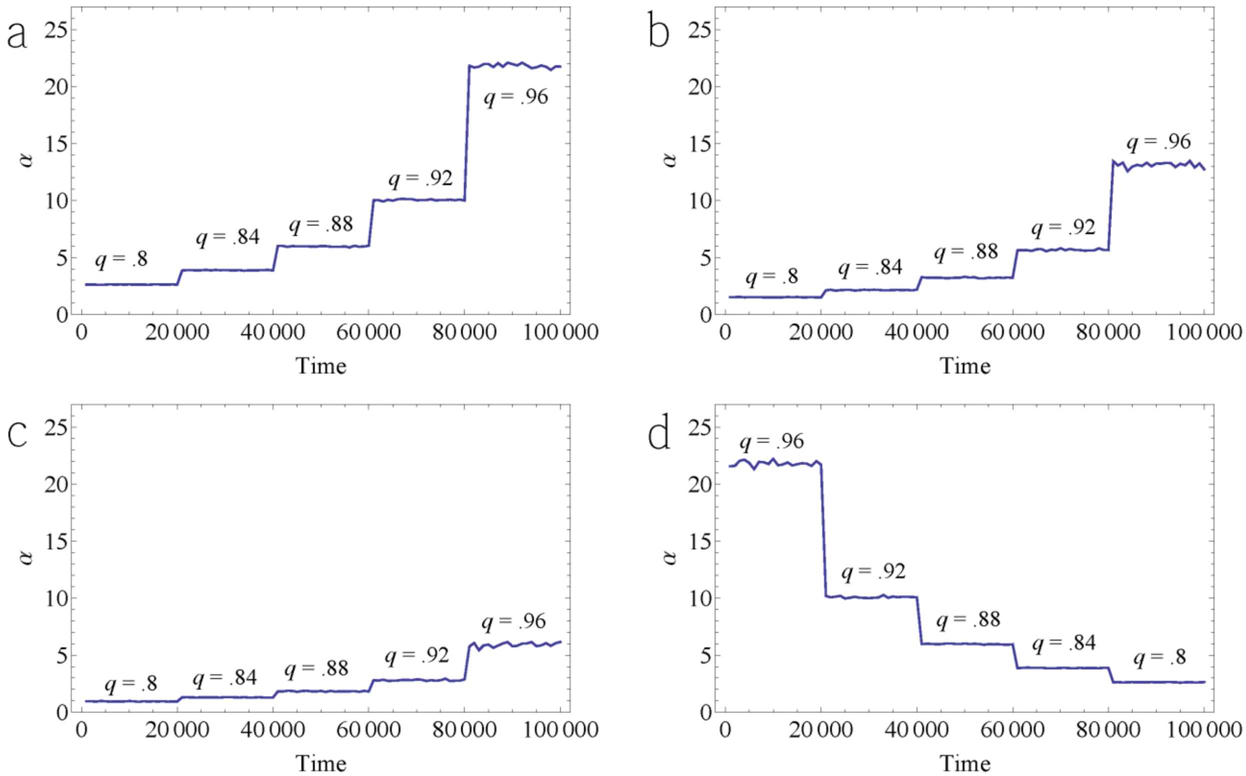

Figure 12: Equilibrium average α responding to change events in transmission fidelity q. Parameters used: In (a) s = .9, (b) s = .6, (c) s = .3, (d) s = .9. In all, N = *30*, simulations run over 100,000 updates, data points averaged over 1,000 time steps.

In Figure 12:a-d we see that the response to changes in imitation fidelity norm q is highly similar to that reported by Andersson (2011) and Andersson (2013); the population responds by rapidly moving to a new equilibrium level determined by the value of q. We also see that skill bias s has no qualitative, but a strong quantitative, impact on the evolutionary dynamics. The population also responds symmetrically if imitation fidelity q is decreased rather than increased (Figure 12:d).

In Figure 13:a-d it emerges clearly that changes in population size N have little effect on skill complexity α unless skill bias s is sufficiently high; in Fig. (13b) with s = .6 even a slight decrease in performance for larger populations can be observed (see also Figure 4), and in Figure 13:c, using s = .3, the only effect of increasing N is a reduction in fluctuations. In Figure 13:d we see that the population responds symmetrically also to reductions in population.

Response to increases (Figure 14:a-c) and decreases (Figure 14:d) in skill bias s are more distinct than changes in N but less so than changes in q.

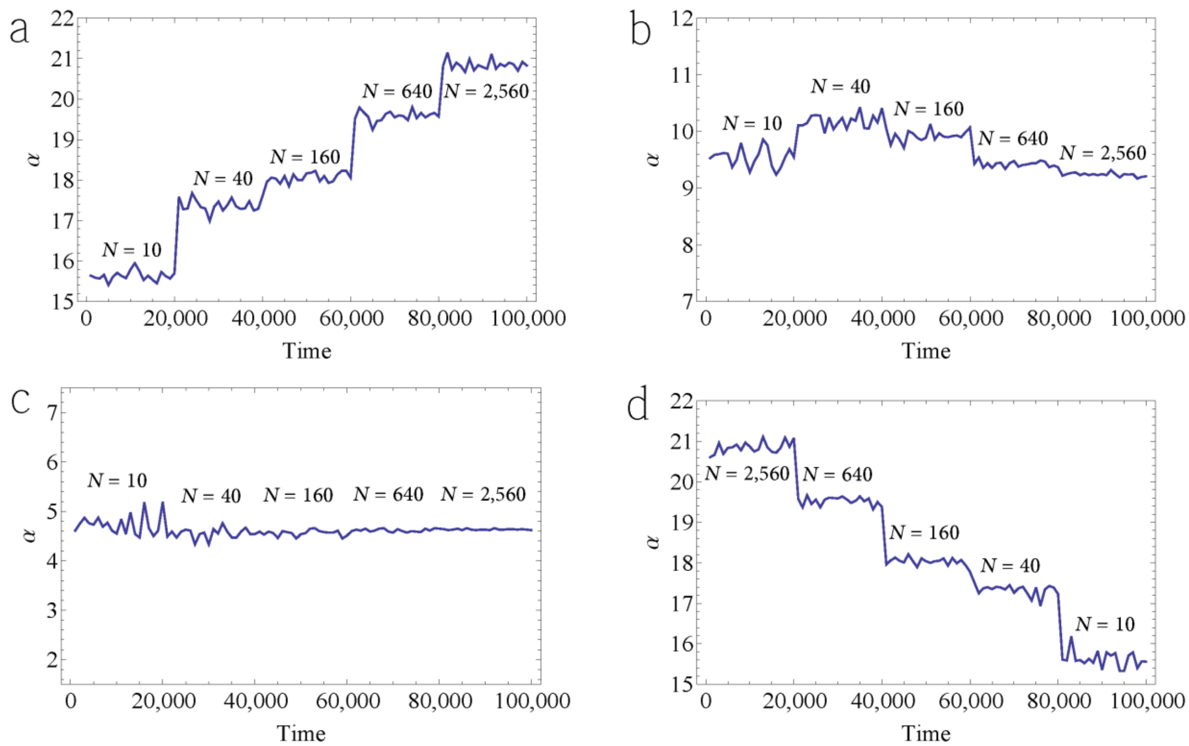

Figure 13: Equilibrium average α responding to change events in population size N. Parameters used: In (a) s = .9, (b) s = .6, (c) s = .3, (d) s = .9. In all, q = .95, simulations run over 100,000 updates, data points averaged over 1,000 time steps.

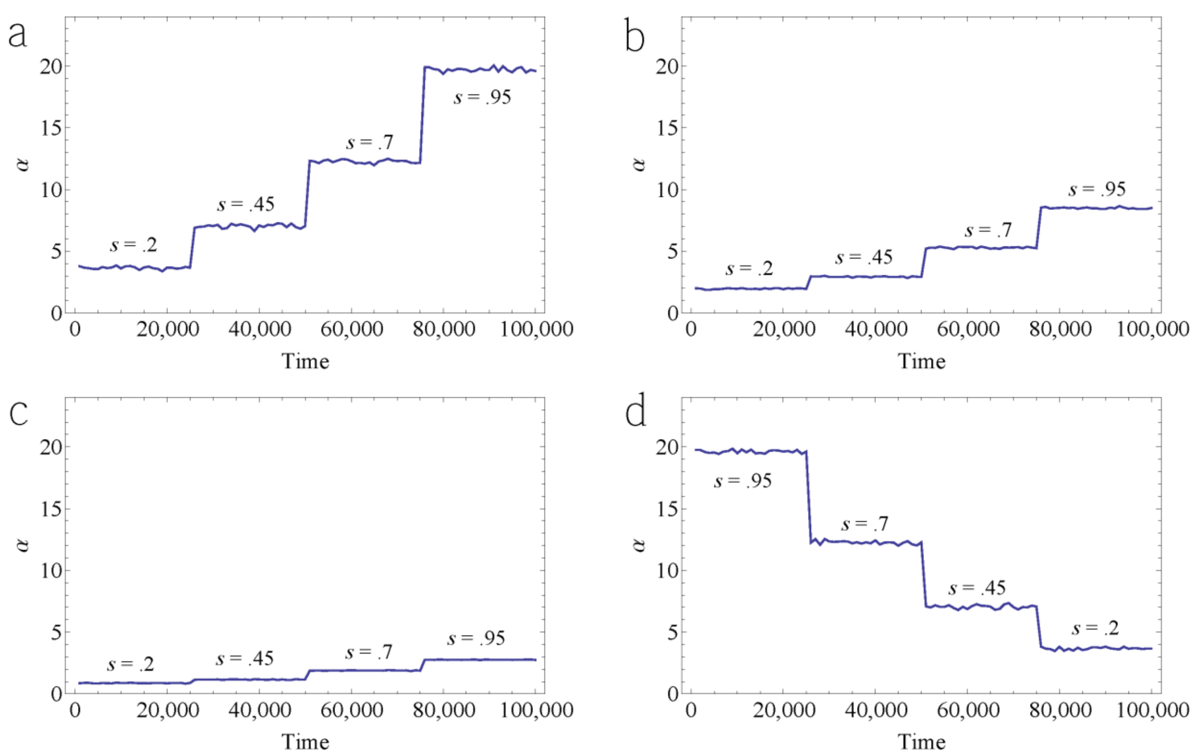

Figure 14: Equilibrium average α responding to change events in skill bias strength s. Parameters used: In (a) q = .95, (b) q = .9, (c) q = .8, (d) q = .95. In all, N = 30, simulations run over 100,000 updates, data points averaged over 1,000 time steps.

ODD (Overview, Design concepts, Details) description of the model

The model description follows the ODD (Overview, Design concepts, Details) protocol for describing individual- and agent-based models (Grimm et al. 2010).

1. Purpose

The basic purpose is to abstractly simulate the emergence of phenomena of interest in cultural evolution on the level of populations of individuals. In particular the model aims to explore factors that enable, stimulate, inhibit or systematically shape the evolutionary cumulativity of cultural knowledge.

The model synthesizes two models from the literature: **(i)** the "Treadmill Model" (Henrich 2004) and **(ii)** the "Glass Ceiling Model" (Andersson 2011, 2013) with the purpose of placing their basic processes – which they propose are important for understanding the issues mentioned above – in the context of one another. This is done by exploring and describing the behavior of the resulting model and comparing it with the two component models as well as with empirical and experimental data and facts.

The immediate purpose then is to see whether this synthetic model does a better job at explaining empirical and experimental data that are in disagreement with the models that are combined individually.

A more overarching purpose is also to thereby demonstrate the need for combining models of single factors that are typically explored scientifically on their own, disregarding to a large extent the potential for emergent behavior when they are combined.

2. Entities, state variables, and scales

The model contains two types of entities: individuals and populations. Individuals are organized into populations and there is only one population. They represent two levels of organization, disregarding intermediate levels for simplicity.

<u>On the level of the population, the state variables are:</u>

N : The number of individuals in the population.

s : The skill bias strength parameter.

q : The imitation fidelity norm

IQ : The IQ norm; always set to IQ = 100, value is arbitrary

$IQsd$: The standard deviation of IQ scores in individuals; always set to IQsd = 15.

α_0 : Initial value of α.

<u>On the level of individuals, the state variables are:</u>

q_i : Imitation fidelity

α_i : Complexity of skill held by the individual

Populations are updated synchronously (generations) simultaneously replacing all individuals. Individuals have generated *IQ* values, but, used only for generating q_i they are not stored.

The time scale is that of "generations", which is not strictly mapped to an absolute time scale.

3. Process overview and scheduling

The model simulates a specified number of *generations*. A generation consists in the execution of the following operations:

1. A new generation of "naïve learners" is generated. For each individual *i*:
    a. Generate *IQ* score with mean IQ and standard deviation *IQsd*
    b. Generate q_i (from *IQ* score and *q*; see Equation 12 in Appendix C in Supplementary Materials)
2. The enculturated individuals *i* in the old generation all receive selection weights according to the scoring algorithm, using α_i and *s*. (see the *skill bias* submodel below and Equation 13 in Appendix C in Supplementary Materials).
3. Each naïve learner *i* is paired up with a "role model" *r* from the old generation. The probability of selecting a specific role model is the selection weight relative to the sum of the selection weights over the old generation.
4. The naïve learners imitate their role model *r*. This gives them a value for α_i, generated from α_r and q_i (see *imitation* submodel below and Equations 5 & 6), using $\alpha_i = 0$ if the generated $\alpha_i < 0$.
5. The learners are no longer "naïve" and become the "old generation"
6. Generation is complete, next generation begins over at #1.

This scheme keeps the population size fixed at *N*.

All population-level parameters can be altered runtime (in practice, *q*, *s* and *N* are). Varying *q* and *s* is straightforward. Altering *q* and *s* is straightforward, and so is altering *N* since we implement no limit on how many learners that can be associated with each role model.

4. Design concepts

<u>Emergence.</u>

We are particularly interested in the qualitative patterns and phenomena that are generated by hypothetical processes on the level of individuals interacting in an evolutionary population. In particular we are interested in increases, decreases and equilibria in α, in cases where the processes controlled by the parameters that we vary (q, s, N) interact in non-trivial ways.

We also link emergent patterns and phenomena that are readily quantified in terms of the model output (primarily α) to more qualitative descriptions of emergent features of the dynamics. For example in Figure 3 and the Supplemental Movies where the action of the processes controlled by the parameters are described.

<u>Adaptation.</u>

Naïve role models have the capability to choose role models based on their performance. Whether this happens using their own cognitive faculties, other individuals or cultural institutions is not specified.

<u>Objectives.</u>

The objective of the naïve individuals – or of the group as a whole – is thereby to attain high values of α.

<u>Interaction.</u>

Interactions are assumed to be sufficiently intensive that role model selection and imitation can take place. As discussed in the article, this may in important cases (most notable in large populations) be a strong assumption.

<u>Stochasticity.</u>

Stochasticity is fundamental to the workings of the model. *IQ* values, and thereby the values of q_i associated with individuals, as well as error terms in imitation are stochastic.

IQ is drawn from a standard distribution which is a standard choice.

Error terms are also drawn from a standard distribution, and this is a stronger assumption since the empirical background is scant and since it is reasonable that different types of skills would generate different patterns of error.

However, since the important feature of stochasticity here is simply that there exists a scatter of values, the results are unlikely to be highly sensitive to the choice of distributions as long as this condition is met.

<u>Observation.</u>

Data can be generated freely from the state variables across a run. There are two main classes of data generated:

1. Time series
2. One parameter expressed as a function of another parameter

Time series are produced by specifying a *sampling frequency* in the unit of generation updates. Averages are calculated across time slots of the specified size and used as representative values for that time slot in visualization.

When a parameter is expressed as a function of another parameter we perform parameter sweeps over one or two spans of parameters values, keeping the rest of the parameters fixed. We specify an initial value, a step size and a number of steps for the parameters to be swept. We average each data point over a specified number of runs with identical parameters settings, providing error bars to indicate how variable the output is. For example, in Figure 4, we investigate equilibrium values of average α in the population as a function of used population sizes.

5. Initialization

The model is initialized by generating a new population as described in Section 3 above, but setting $\alpha_i = \alpha_0$. Due to the formulation of the model, its behavior is highly robust to the choice of $\alpha_0 > 0$. We arbitrarily use $\alpha_0 = 4$.

7. Submodels

We have the following two important submodels:

<u>Skill bias</u>

The skill bias submodel is technically described in the Article (Equation 13). Its key role is to provide a parameter *s* that allows us to gradually tune the selection regime between pure drift ($s = 0$) and strictly selecting only the best role model ($s \to 1$). The function (Equation 13) is undefined for $s = 1$, but in finite populations, a value of *s* sufficiently close to unit yields the sought for limit behavior.

The model, it should be added, is not strictly *only* of skill bias, but of any bias that affects the quality of the role model with which a naïve learner will learn from (e.g. prestige bias may also have such an effect). Skill bias is typically seen as involving a direct assessment of the skill of potential role models whereas prestige bias cuts across domains. For example, that we may gain a more general trust for accomplished individuals regardless of what they are accomplished in.

<u>Imitation and innovation</u>

Imitation and imitation is conflated into a single model of transmission. The model is described technically in the Article and is closely inspired by the imitation model introduced by Henrich's (2004; see Article bibliography), but the fixed average error α is replaced with a function of α and *q*, which effects the synthesis with the Glass Ceiling Model by Andersson (2011, 2013).

References Cited

Andersson, Claes
2011 Palaeolithic Punctuations and Equilibria: Did Retention Rather Than Invention Limit Technological Evolution? *PalaeoAnthropology* 243–259.

Andersson, Claes
2013 Fidelity and the Emergence of Stable and Cumulative Sociotechnical Systems. *PalaeoAnthropology* 88–103.

Flynn, James R.
1987 Massive IQ Gains in 14 Nations: What IQ Tests Really Measure. *Psychological Bulletin* 101(2):171–191.

Grimm, Volker, Uta Berger, Donald L. DeAngelis, J. Gary Polhill, Jarl Giske, and Steven F. Railsback
2010 The ODD protocol: A Review and First Update. *Ecological Modelling*, 221: 2760–2768.

Henrich, Joseph
2004 Demography and Cultural Evolution: How Adaptive Cultural Processes Can Produce Maladaptive Losses: the Tasmanian case. *American Antiquity* 69(2):197–214.

Price, George R.
1970 Selection and Covariance. *Nature* 227:520–521.

Vaesen, Krist
2012 The Cognitive Bases of Human Tool Use. *Behavioral and Brain Sciences* 35:203–218.