Journal of Evolutionary Psychology, 2011, 125-141
DOI: 10.1556/JEP.9.2011.18.1

THE SELECTIVITY OF SOCIAL LEARNING AND THE TEMPO OF CULTURAL EVOLUTION

R. ALEXANDER BENTLEY[1,*] AND MICHAEL J. O'BRIEN[2]

[1]Department of Anthropology, Durham University
[2]Department of Anthropology, University of Missouri

Abstract. Many modern studies of cultural innovation and demographic change rest on the proposition that social learning is a key process in the spread of novel variants. We agree with this proposition, but we also suggest that the selectivity of social learning, with respect to a skill or knowledge, has largely been overlooked in considering how the tempo of cultural evolution depends on population size. We evaluate contrasting predictions ranging from cases of extreme selectivity, where everyone learns from the best individual in the group, to entirely nonselective, unbiased social learning, where everyone learns from one another. At the highly selective end of the spectrum, population size should correlate strongly with the tempo of cultural evolution. At the nonselective end, where social learning is unbiased, there should be little correlation between population size and tempo of cultural evolution. For any given case study, characterizing the relative selectivity of social learning is crucial to successfully predicting the effect of population size on cultural evolution.

Keywords: extreme value theory; "Tasmania effect"; social learning; tipping points

INTRODUCTION

Hardly anything that humans learn is void of social interaction. With collaboration and learning from others, science, for example, is certainly a social activity (HULL 2001; GUIMERÁ et al. 2005; O'BRIEN, LYMAN and SCHIFFER 2005). We all know that science is a collective endeavor and that the more scientists there are, the more knowledge is built on the shoulders of previous scientists. In the social sciences, the more general idea of dispersed knowledge in human populations is at least as old as the work of sociologist Emile DURKHEIM (1898) and is plainly evident in the work of economist Friedrich HAYEK (1949) and his notion of "catallaxy," or the self-generating economic order arising from actors observing and learning from others. With daily life on the Internet no doubt reminding us, the "collective brain," or "distributed mind," concept has made a resurgence as an explanation for what drives economic and cultural complexity (RIDLEY 2010), particularly in studies of cultural transitions in prehistory (e.g., HENRICH 2004; POWELL, SHENNAN and

[*]Corresponding author: R. A. BENTLEY, Department of Anthropology, Durham University, Dawson Building, South Road, DH1 3LE, Durham, United Kingdom; E-mail: r.a.bentley@durham.ac.uk

THOMAS 2009; RICHERSON, BOYD and BETTINGER 2009; ZILHÃO et al. 2010). Ideas of group size are central to evolutionary psychology, with one central theme being DUNBAR's (1993) proposed limit to the number of people – roughly 150 – with whom one can maintain stable social relationships. There is an upper limit to population size, at which point information must be stored in other media besides human minds.

The phenomenon of a collective brain falls most naturally into the realm of sociology and economics. It is therefore less well studied in anthropology or evolutionary psychology, where it can often usefully be overlooked through simplifying assumptions about one of the three evolutionary processes – variation, transmission, and selection. Evolutionary psychology usually emphasizes natural selection on Plio-Pleistocene hominins, which genetically shaped the phenotypic plasticity that humans now use to adapt to modern life. Focusing on selection, so that the other two processes need not be finely detailed, works well where one behavior is clearly better than another one and therefore can be predicted to be selected over time (e.g., making spears with a sharp rather than a blunt point or foraging for food efficiently rather than inefficiently).

By contrast, studies of cumulative cultural evolution focus on transmission, through human interaction and social learning, which is posited not only as the foundation of cultural complexity and variation but as being integral to hominid evolution (BOYD and RICHERSON 1985, 2005; HRDY 2009). For much of prehistory, accumulated knowledge was passed down, with occasional variation, from generation to generation. Through a positive feedback, presumably between specialization and accumulation, human knowledge has intensified exponentially in the millennia since the Neolithic. The total diversity of human material culture, however one might grossly quantify it, has increased perhaps eight or nine orders of magnitude since the Paleolithic (e.g., BEINHOCKER 2006; RIDLEY 2010).

What makes human cultural evolution unique is an extraordinary social learning ability, which evolved under evolutionary selection and has considerable consequences for evolutionary *transmission*. Multiple animal species are able to learn, and a good number of them on occasion practice social learning (GALEF and LALAND 2005; LALAND and READER 2010), but humans are more accurate and complex social imitators than any other animal. Humans continue to "learn things from others, improve those things, transmit them to the next generation, where they are improved again, and so on," and this process continues to lead to the "rapid *cultural* evolution of superbly designed adaptations to particular environments" (BOYD and RICHERSON 2005:4, emphasis in original).

Social learning, a necessary prerequisite to cultural evolution, is a process of transmission of knowledge between minds, and there is a large literature on the various routes of information transmission between individuals as well as on the biases in choosing these routes, such as prestige bias, conformity bias, and so on. If we are concerned with population-level effects of this transmission over many human generations, it is useful to consider prehistoric knowledge as being stored and transmitted among people. This concept is increasingly commonplace in our modern, wiki-media era of computers and information storage, and we are seeing more research into how humans have stored and retrieved information in other people, and subsequently in cave paintings, writing, built environments, and material culture (e.g., RENFREW and SCARRE 1998; BOLENDER 2007; GABORA 2008; POWELL, SHENNAN and THOMAS 2009).

Through this distributed-mind approach, one can equate collective knowledge capacity with population size, that is, the number of people communicating and storing their specialized ideas. This has had philosophical and empirical ramifications. Philosophically, it suggests that population size has a direct bearing on thought because what goes through an individual's mind is largely derived from other minds. Thus an individual's thoughts are but a sample of what is being thought around that person. Empirically, it means that the tempo of cultural change – partly a function of how much cultural information is preserved and passed on – may depend on population size (SHENNAN 2000). An exciting new body of work posits that increasing population size rather than changes in biological ability may underlie some major shifts in cultural prehistory, such as the Upper Paleolithic revolution (POWELL, SHENNAN and THOMAS 2009) or the much-debated cultural capacities of late Neanderthals (ZILHÃO et al. 2010). Conversely, population bottlenecks are now proposed to reduce cultural knowledge collectively, as HENRICH (2004) argued for prehistoric Tasmania.

Somewhat hidden within this discussion of human behavior as a collective phenomenon are quite specific assumptions about social learning that directly affect the predictions for how population size correlates with the pace of cultural evolution. Here we consider how underlying assumptions about the selectivity of social learning are crucial to the outcome of models of how knowledge accumulates in recognizable patterns over time. This selectivity applies to the social-learning strategies people followed, for which LALAND (2004) presents a useful list:

- Copy the majority
- Copy successful individuals
- Copy if better
- Copy good social learners
- Copy kin
- Copy friends
- Copy older individuals

Some of these strategies are more selective, as they refer to skill level as the main criterion (copy better, copy good social learners, copy successful), whereas others are focused on social criteria (copy the majority, copy kin or friends, copy older individuals). One additional strategy, "copy if rare," is, in terms of variant frequencies rather than psychology, only a small step away from independent invention and could be considered a form of preserving variation.

Here we contend that there is a crucial difference in the effects of copying based on selection for the knowledge or skill level as opposed to copying based on social criteria. For clarity we consider this as a spectrum of selectivity with respect to the action itself, ranging from strong selectivity, such that the "best" version is selected by each learner from within the group, to weak or nonexistent selectivity, where social learning is governed by factors not directly related to the activity itself. Where social learning lies on a continuum between these end points is important for understanding how the tempo of cumulative culture evolution corresponds to the capacity for knowledge storage. In order to make this argument, we need to briefly review theories that focus strongly on one extreme or the other, particularly HENRICH's (2004) model, which we see as based on extreme selectiveness of social learning. The model can yield surprising insights whenever the assumption is appropriate, but it seems in danger of becoming widely applied without explicit consideration of this assumption. We discuss HENRICH's (2004) model first in order to show that the selectivity of social learning is an important factor in the predicted effect that population size has on the tempo of cultural evolution. We then show that if social learning is largely nonselective, or unbiased, then population size actually has minimal effect in this regard.

STRONG SELECTIVITY

If people bias copying toward selection of skill or knowledge, then the tempo of cumulative cultural evolution can become strongly related to population size. How population size affects this tempo depends on the assumptions that are made about the strictness of the selection and the range of variation upon which selection is acting. Here we explore how different assumptions would affect cumulative evolution. When these assumptions are varied, we expect entirely different effects for different combinations of parameter ranges. In terms of the strength of selection, an extreme but useful assumption can be that the most skilled living individual is always the model for learning, as when a group of novices learns from a grandmaster. In an important paper, HENRICH (2004) showed how skills that are learned from grandmasters might be passed on, and accumulated, from one generation to the next. The model predicts a critical minimum size for a population to accumulate skill or knowledge over the generations. Using a case study from prehistoric Tasmania, Henrich suggests that a population bottleneck in the prehistory of the island would have dropped it below this minimum threshold and hence led to a retrogression in the skill set, including both knowledge and technique, of the Tasmanian cultural repertoire. Especially susceptible were complex skills–"tools that are hard to learn to make, and easy to screw up" (HENRICH 2006:776).

In this model, skills advance through exceptional, gifted naturals rather than through well-schooled regular Joes. This extreme selectivity is the reason Henrich's model is conservative with respect to the deteriorating effects of reducing population size on cumulative adaptive evolution: If the assumption of extreme selectivity were relaxed, the predicted effects would just happen more readily. The conservatism of the model with respect to shrinking populations, however, becomes *just the opposite* with respect to increasing populations, as we discuss in the following section. In order to see where misunderstanding might creep in, we need to look at the model in some mathematical detail.

HENRICH's (2004) model is an outgrowth of a larger body of theory that deals with extreme events, known as Extreme Value Theory (EVT). EVT was originally developed to account for the occurrences of extreme events in nature, such as floods and landslides, which can be great shapers of events, even though they occur rarely and not within a normal (Gaussian) probability distribution (TURCOTTE 1997). A single summer flood, for example, can move massive boulders downstream in a single day that otherwise would remain at rest for decades or centuries. The EVT describes the distribution of observed values x from the following cumulative distribution function (CDF):

$$H_{\xi,\mu,\beta}(x) = e^{\frac{\xi(\mu-x)}{\beta}-1} \tag{1}$$

where μ is the mode and β is the standard deviation. The parameter ξ ($\neq 0$) controls for the "fatness" of tail of the distribution, which is crucial to the outcome of the EVT (FRANK 2009). In the special case where ξ is zero, the EVT predicts the CDF to be

$$H_{\mu,\beta}(x) = e^{-e^{\frac{\mu-x}{\beta}}}. \tag{2}$$

The CDF in equation (2), evaluated at x, gives the probability of choosing a value equal to or less than x. Hence when x is large, $H(x)$ converges on 1.

HENRICH (2004) innovatively adapted EVT for social learning by assuming that in any given group, individuals will select the most extreme person, that is, the most skilled, from whom to learn that skill. Henrich considered a renewable pool of N individuals, representing the size of successive generations or age cohorts of a society. In each time step, all individuals attempt to copy the most skilled individual in the group, whose skill level is denoted by z_h. This can be a fair assumption for many activities (e.g., each hunter emulates the best hunter in his group or an aspiring golf or tennis player gets lessons from the best pro in town). Just as the largest flood events are primarily responsible for moving boulders downstream over the centuries, the most skilled individuals are most responsible for advancing a craft tradition over the generations. As HENRICH (2006) pointed out, this is a conservative assumption regarding the potential loss of information with decreasing overall population size, because if the most skilled individual is less than perfectly accessible, the cumulative knowledge deteriorates faster.

In Henrich's formulation, the skill of the most proficient individual, z_h, is α units better than the mode, μ, of the skill distribution, i.e., $\mu = z_h - \alpha$. With equation (2) being the CDF, the corresponding probability distribution function (PDF) is the derivative of equation (2):

$$p(x) = \frac{\partial H}{\partial x} = \frac{1}{\beta} e^{\frac{\mu - x}{\beta}} e^{-e^{\frac{\mu - x}{\beta}}}. \tag{3}$$

This PDF is a particular form of the double exponential, or Gumbel, distribution (e.g., FRANK 2009), which in this case yields the probability that a particular imitator will wind up with skill level *x*.

In Henrich's depiction, shown in *Figure 1,* the fatness, or spread, of the Gumbel PDF is defined by β, and the position of the mode, μ, is determined by z_h and α (given that $\mu = z_h - \alpha$). The loss value, α, is always a setback, but the β value describes the variation from that setback in either direction. Note that for large values of the skill level x, where $x > \mu$, the Gumbel distribution (equation 3) behaves essentially like a declining exponential, $\sim e^{-x/\beta}$. Hence at the high-skill end, the skill of the best imitator declines exponentially in probability.

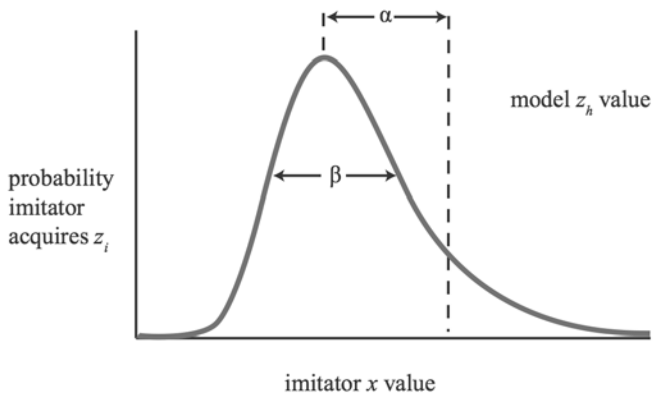

Figure 1. The Gumbel distribution for imperfect imitation, as depicted by HENRICH (2004)

Because the mode of the Gumbel distribution can increase over successive generations through the consecutive selection of extreme values, what HENRICH (2004) and POWELL, SHENNAN and THOMAS (2009) refer to as "cumulative adaptive evolution," is essentially a simple model of skills represented by the distribution, feeding back on themselves through intergenerational learning and becoming more specialized as the accumulated skill level increases. Henrich showed that the ratio of α to β, the average loss to the width of the Gumbel distribution, is the crucial variable, with opposing effects on adaptive evolution: "α operates against adaptive evolution, while β, the tendency of individuals to make *different* inferences from observing the same thing, *favors* adaptive evolution. The more individuals tend to make different inferences, the faster cultural evolution goes – or the more likely it is to be adaptive" (HENRICH 2004:202, emphasis in original).

In other words, by ratcheting the most skilled individual as each generation's model, successive generations sampling from these Gumbel probability distributions would gain skills over time. This is a standard prediction of cumulative cultural evolution (TOMASELLO, KRUGER and RATNER 1993; BOYD and RICHERSON 2005), but in this case the demographic conditions under which it occurs are very specific. The ratchet works if we expect in each generation a sample where the forward error of at least one individual is larger than the setback, α. The chance of an imitator obtaining a skill level greater than the best of his generation, z_h, involves the CDF in equation (2), such that the chance is $1 - H(z_h)$, or

$$P(x > z_h) = 1 - e^{-e^{\frac{\mu - z_h}{\beta}}} . \tag{4}$$

Because $\alpha = z_h - \mu$, this can be written as

$$P(x > z_h) = 1 - e^{-e^{\frac{-\alpha}{\beta}}}. \tag{5}$$

For reasonably large α/β, the exponent, $-e^{-\alpha/\beta}$, is small, so we can make a Taylor series approximation[1] that yields

$$P(x > z_h) \approx e^{\frac{-\alpha}{\beta}} . \tag{6}$$

Hence, the chance of a given imitator becoming better than the best of his generation declines exponentially with the distribution ratio α/β.

Here is where population size enters the picture with respect to the extreme-selectivity model. For any given α/β ratio, the larger the population, N, the better the chance of there being at least one individual in each generation who surpasses the skill of the best individual in the previous generation. As shown in *Figure 2,* Henrich demonstrated that this minimum population size, N^*, for cumulative adaptive evolution increases exponentially with α/β:

$$N^* = 0.56e^{\frac{\alpha}{\beta}} . \tag{7}$$

Equation (7) thus defines a population "tipping point," N^*, where cumulative adaptive evolution becomes possible. This is the essence of a particular demographic model for the Upper Paleolithic (POWELL, SHENNAN and THOMAS 2009), and HENRICH (2004) also suggested other possible examples from small-scale societies. If humans have always lived in groups of at least a few tens of members, then the interesting effects are for N^* values larger than that (otherwise the model would not explain any transitions). According to equation (7), this requires an α/β ratio of at least 4 or 5 (HENRICH 2004, 2006); N^* is about 30 at $\alpha/\beta = 4$, and N^* reaches 100 when α/β is about 5.2. Referring back to *Figure 1,* however, Henrich's depiction of the Gumbel distribution uses an α/β ratio of only about 1. In other words, *Figure 1* does not properly represent the condition for cumulative cultural evolution.

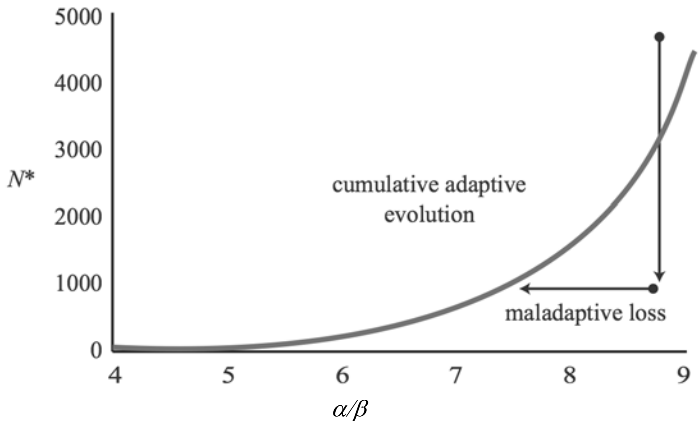

Figure 2. Threshold values of N^* versus α/β for the HENRICH (2004) model, showing predicted regimes of cumulative adaptation and maladaptive loss

If we look instead at the Gumbel distribution for the minimum α/β of 4 or 5, it looks quite different (*Figure 3*). We see that in this case the most skilled individual in the population is well off to the right, essentially several widths of the distribution away from the majority. In *Figure 3,* the teacher is highly exceptional compared to the majority of the population. The person is so far out of the normal range that no one in the majority could fail to recognize the teacher as anything but a prodigy.

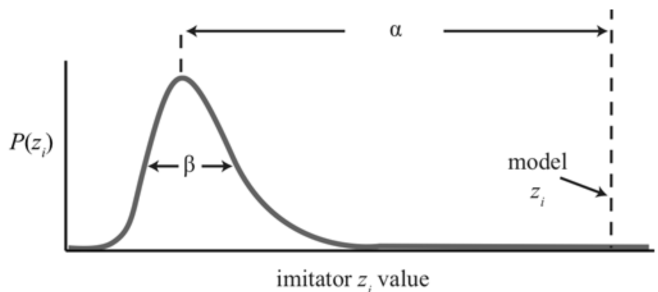

Figure 3. Gumbel distribution as in *Figure 1* but with $\alpha/\beta = 5$

In summary, the prediction of a demographic tipping point in models of extreme selectivity (e.g., HENRICH 2004; POWELL, SHENNAN and THOMAS 2009) incorporates two strong assumptions: (1) that the tail of the skill distribution decreases exponentially, converging to a Gumbel distribution over multiple generations, under which there is a *very* exceptional teacher ($\alpha/\beta > 4$) – a true genius who sets the bar for the next generation; and (2) all learners copy the most skilled teacher. The exponential tail of Gumbel distribution (or any other exponentially tailed distribution; see FRANK 2009) is essentially the reason why the minimum population for cumulative adaptive evolution, N^*, increases exponentially with $\langle\alpha/\beta$ (*Figure 2*).

WEAK SELECTIVITY

In extreme-value theory, many different initial distributions attract to the Gumbel distribution over successive generations of selecting the maximum value. Other distributions are possible in EVT, however, including the Fréchet distribution from successive selection from a power-law distribution and a Weibull distribution, where there is a finite limit to the tail of the skill distribution (FRANK 2009).

Of course, in social phenomena the normal (Gaussian) distribution is quite common, as the name implies. Selecting the maximum value from a normal distribution over successive generations converges toward the Gumbel, but because the normal is still so often observed, we suspect that the selection is not actually so extreme. Skills in mathematics and science, for example, are normally distributed (HYDE and LINN 2009). As shown in *Figure 3,* given that the tail of the normal distribution declines much faster than that of the Gumbel distribution, "geniuses" that could be expected under the Gumbel distribution are highly unlikely under a normal distribution.

Consider a hypothetical, but numerically explicit, example used by HENRICH (2006:775): "Imagine a group of 1,000 novice archers trained by a grandmaster, who himself scores 150 on the required archery target examination. After training, most of these newly fledged archers score between 20 and 40 on the examination." Let us assume this example represents the threshold of cumulative adaptive evolution, i.e., $N^* = 1000$. Using equation (7), this means that $\alpha/\beta = 7.5$ and that the grandmaster, with score 150, is probably over seven and a half widths ($7.5^*\beta$) better than the mode of the students' distribution, which we can assume is about 30. So β is about 16 (120/7.5). Now, what if the distribution were normal instead of Gumbel? To get a sense of it, let the width, β, be the standard deviation of this normal distribution and let the mode of 30 be its mean. Under a normal distribution with mean = 30 and standard deviation = 16, about 99% of the archers will be within three standard deviations – that is, they will have scores less than 78 – and about 99.9% will have scores less than 82. The chance of finding the grandmaster's value of 150 is effectively zero (less than one in a trillion). So if the learners' abilities are normally distributed, it is virtually certain that no one in the next generation will even be close to the grandmaster.

To illustrate the point further, consider a less extreme degree of selectivity in social learning, where we use Gaussian distributions rather than the Gumbel distribution. There is a substantial literature on cumulative culture evolution, and our point here is to illustrate how relaxing the assumption of extreme selectivity can easily yield different results. For example, consider N learners in each generation, as HENRICH (2004) proposed, who copy their skills from the previous generation, but instead of everyone copying the absolute best with some error, we have each individual copying the *average* skill level (averaged across individuals in the group) plus some minor learning error that is normally distributed around zero (positive or negative). This standard model yields a random walk in terms of the mean skill level for the group (*Figure 4a*), with stochastic change that can go up *or down* over time. Also, such a model is unpredictable, in that each random walk is unique. Copying the majority, then, where behavior is continually drawn to the *status quo*, could, under these assumptions, make cumulative adaptive evolution merely a matter of drift (HAMILTON and BUCHANAN 2009).

In other words, the nature of the dependence on N, assumed in a growing number of studies on cumulative culture evolution, appears to require an extreme selectivity for the skill itself as well as an exponential tail of the skill distribution. Many equally plausible models do not have this effect. In a random walk, for example, runs for, say, $N = 10$ would be quite similar in terms of variation and change to runs for $N = 100$ (*Figure 4b*). This is true even if one assumes the error is always positive, to represent perhaps "what most scientists do most of the time," as ANDERSON and ABRAHAMS (2009:1515) put it. If each individual makes a small but positive *improvement* of some random amount (drawn from a Gaussian distribution), there would, of course, be continual improvement in average skill level over time but still little effect of the population size, N, on the rate of improvement.

Finally, in the extreme of social interaction without any selection for skill or knowledge, the overall dynamics can become equivalent to the neutral model of culture evolution (NEIMAN 1995), where the value of the skill itself is neutral with respect to its chance of being copied. On the population scale, much of fashion is like this, such that fashion in large populations can effectively be modeled *as if* people were copying one another at random (HAHN and BENTLEY 2003; BENTLEY, HAHN and SHENNAN 2004; HERZOG, BENTLEY and HAHN 2004; BENTLEY, ORMEROD and BATTY 2010), just as the Ideal Gas Law predicts pressure *as if* the molecules were moving about randomly. Skill-level distribution is for the most part irrelevant in the cases of easily learned activities such as using certain fashionable names and words (HAHN and BENTLEY 2003; BENTLEY 2008; BERGER and LE MENS 2009), listening to certain music (SALGANIK, DODDS and WATTS 2006), and decorating with certain styles (NEIMAN 1995; BENTLEY and SHENNAN 2003).

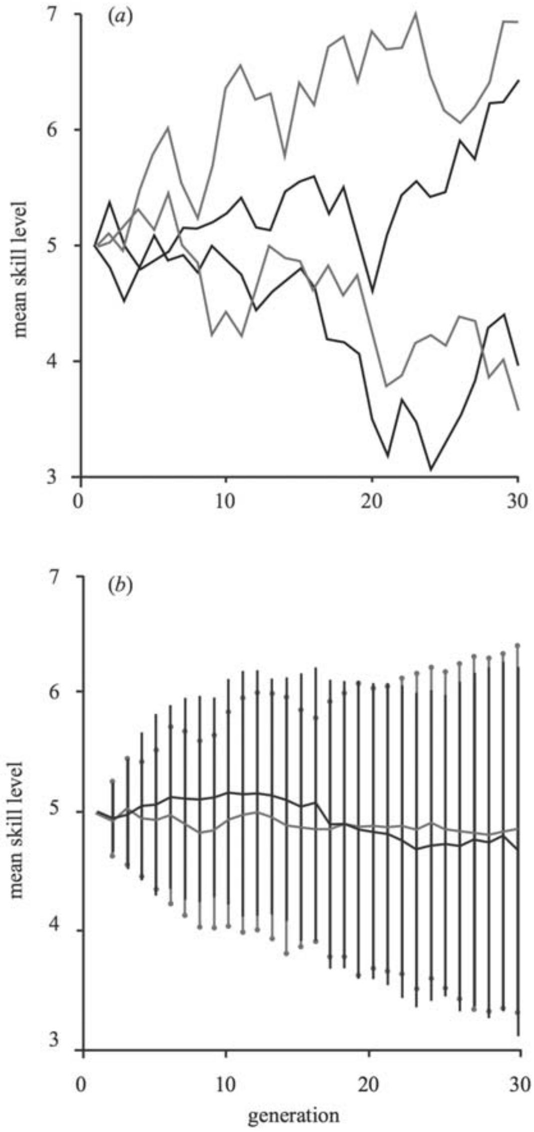

Figure 4. Characteristics of a simple learning model where learners copy the average of normally distributed skill level plus or minus and error term: (a) mean skill (or knowledge) level versus generation number for several representative runs, one with $N = 10$ individuals (gray line) and the other with $N = 100$ individuals (black line); (b) results averaged over 30 runs for $N = 10$ (gray) and $N = 100$ (black). The error bars show one standard deviation for the 30 runs, which are nearly identical for $N = 10$ (gray bars with filled circles at the ends) and $N = 100$ (black bars). As expected for a random walk, the standard deviation increases as approximately the square root of the number of generations ($r^2 > 0.95$ for both $N = 10$ and $N = 100$)

The neutral model applies best to behaviors that are not essential to survival (hence the term "neutral"), where what is chosen has no *inherent* value relative to other options – no inherent "quality" is assigned to any of the choices. This model reproduces certain emergent patterns in collective behavior that can be used as a background against which selection for quality becomes evident. These patterns include (a) long-tailed distributions of popularity; (b) continual turnover, with any new variant having a finite (if usually small) chance of becoming highly popular; and (c) unpredictability, in that events are never the same from one time step to the next. Through different combinations of its few parameters (population size, innovation, and memory), the model can be used to predict change rates as well as, at least potentially, to distinguish copying from other forms of collective behavior.

The outcomes of the neutral model include the much discussed long-tailed, including power-law, distribution as an emergent property. Previous models for long-tailed distributions, dating back to the early 1900s, tend to be pre-programmed with strict growth by proportionate advantage, which have been unsatisfying as an *explanation* for long-tailed phenomena, because those with less can never overtake those with more. In the neutral model, there inevitably emerge a few choices that are overwhelmingly more popular than the majority, but these most popular choices are not necessarily any better than any others. This is a reason why copying the popular or copying the (perceived) majority is, in several statistical respects, not much different than neutral copying (BENTLEY and SHENNAN 2003; MESOUDI and LYCETT 2009). In each case, the more popular someone or something is, the more likely the person or object is to be copied. A bias toward copying the popular just makes the effect more extreme, with the limit being everyone copying the same thing or person.

Whereas these distributions remain stable in form over time, they nonetheless undergo continual turnover in composition of individual components. The rise and fall of fortunes within the long tail are seen in fashion popularity and in trendy buzzwords, for example. As a result, constituents of the resulting long-tailed distribution are in continual flux, in just the way that characterizes pop-culture elements in the real world. The turnover results from a balance of innovation, which introduces new ideas, and random drift, by which variation is lost through sampling. New ideas become highly popular by chance alone and then over time become replaced by others, all through drift.

In sharp contrast to the extreme-selectivity model discussed above, as well as other selection models where relative popularities of diffusing ideas eventually sort according to their inherent fitnesses (e.g., KANDLER and LALAND 2009), the neutral model predicts continual turnover, which will increase with the fraction of innovators in the population but (somewhat counterintuitively) not on population size itself – a prediction that is consistent with real-world data (BENTLEY, HAHN and SHENNAN 2004). This is broadly similar to the fact that population size does not affect the pace of fixation under neutral genetic evolution (e.g., GILLESPIE 2004).

DISCUSSION

In summary, regarding the accumulation of specialized knowledge over time, the selectivity of social learning for that knowledge conditions how population size will affect the pace of knowledge/skill accumulation. In the future, a goal is to fill in the spectrum from unbiased social learning at one end, moving carefully toward extreme selection at the other, perhaps through a parameter that represents the variable fitness of different ideas. Increasing this parameter should decrease turnover, increase predictability, and make population size more important. As rational selection is introduced, there may even be a crucial tipping point in the effect of group size on aggregate behavior. In other words, the larger the population, the more ideas that can be accumulated.

Our exploration identifies selection versus interaction as a significant distinction in how knowledge and skill accumulation occurs. We began with the most extreme selection for skill, which assumes (through the Gumbel distribution) that extreme outliers are rare but not impossible occurrences and that individuals can locate, access, and learn from them. These are strong assumptions, which is why HENRICH (2006) rightly described his model as highly conservative when considering the *loss* of accumulated skills through a population bottleneck. HENRICH's (2004) focus on bottlenecks provides new insight, for example, into a fascinating debate surrounding the Pirahã of Brazil, whose language arguably lacks recursion (EVERETT 2005), whose counting system appears rudimentary (GORDON 2004), and whose very experience of the world, EVERETT (2005) argues, is restricted to the present. These conclusions are heavily debated (e.g., LEVINSON 2005), but in the debate only a few researchers, such as BOLENDER (2007), have considered demographic factors explicitly.[2]

When considering *growing* populations, however, and the corresponding *accumulation* of skills, the extreme selectivity model is not conservative. At face value, it predicts that population growth can lead to a tipping point, where knowledge accumulation starts to take off. In this case, it is mainly strong selection (Gumbel distribution) that accounts for the tipping point. If we assume that learning involves some social interaction, and reasonably substitute a normal distribution for the Gumbel distribution, we find that the effect of population size diminishes. Furthermore, under certain other reasonable assumptions, cumulative knowledge can drift up or down, regardless of population size.

The question then becomes, how rare (or common) are such outlier individuals, and is it reasonable to assume that other individuals in the group can locate, access, and learn from them? This question would seem to be central to the issue of the rapid and cumulative buildup of skill and knowledge – the feature that sets human culture apart from that of other animals (BOYD and RICHERSON 1996; GEARY 2007). POWELL and colleagues (2009), for example, used Henrich's model to propose that the explosion of cultural evolution in Europe ca. 40,000 B.C., traditionally considered the signature origin of biologically modern humans, could merely reflect a population increase with no necessary changes in human cognition. They added stochastic and geographic elements to Henrich's model to show how chance clusters of local migrating populations could, by exceeding the crucial population threshold, begin to undergo cumulative cultural evolution over generations. The implications of the findings by Powell and colleagues are quite radical. Decades of explanations for the Upper Paleolithic transition featured cognitive changes in the human mind (e.g., MITHEN 1996; KLEIN 2002) as opposed to demographic changes in human populations (BOLENDER 2007). This may warrant a look at how strongly the model of POWELL and colleagues (2009) depends on the assumptions regarding skill selection.

As LALAND (2004) lists, different possibilities for interaction include copying the majority, copying the most popular or prestigious, or even, we would add, copying virtually at random. Surely these different learning strategies imply substantial differences for knowledge/skill accumulation, which may be directed by social interaction as well as by selection for skills. Particularly pertinent is the form of the skill distribution, which is crucial for predicting the dependence of the skill-selection model on population size.

Looking to the future, there are even more fascinating questions because the selectivity of social learning is changing dramatically online (e.g., KING et al. 2009). On the one hand, wherever people select knowledge online with great discernment (e.g., enabled by a search engine), we could expect a dramatic acceleration in accumulated knowledge as a result of the concomitant increase in effective population size online. Similarly, the progress of science should increase along with the expanding Internet database, provided that online searches for scientific information are strongly selective. On the other hand, as more and more information is stored, social learning becomes less selective. In this century, online social-network media have proliferated so much that researchers are re-adopting the classic view (e.g., DURKHEIM 1898) that people behave as elements of collective social networks rather than as autonomous, self-directed individuals (e.g., BORGATTI et al. 2009; CHRISTAKIS and FOWLER 2009). So much online information is available that academics are constantly re-inventing the wheel in isolated niches, which threatens to retard scientific advancement through more sorting and more clique-based fractionation (GUIMERÀ et al. 2005).

CONCLUSION

A foundational principle of economics, the distribution of knowledge is indisputably among the keys to human society. Among the key determinants of how fast this collective brain grows is the degree of selectivity in social learning. If individuals are selective and accurate in finding the most skilled model for copying, then the pace of cultural evolution depends strongly on population size. If learning is relatively unselective, however, then the tempo of cultural evolution depends only weakly on population size, if at all.

In what appears to be an emerging conventional wisdom equating population size with cultural/technological complexity, there is a danger in overlooking an important assumption of the particular model invoked: that the most skilled individual (or relevant knowledge) in the population is accurately identified. When this assumption is not true, knowledge accumulation actually has little dependence on population size. As a result, an explicit consideration of selectivity of social learning is important for understanding the demographic origins of human technological complexity, past and present.

ACKNOWLEDGMENTS

We thank Tom Dickins for his help, as well as that of two anonymous reviewers. We also thank Melody Galen for drafting the figures. RAB acknowledges the Leverhulme Trust's support of the "Tipping Points" Research Programme at Durham University.

NOTES

[1] The function e^x has the Taylor series expansion

$$e^{-x} = 1 - \frac{x}{1!} + \frac{x^2}{2!} - \frac{x^3}{3!} + \cdots$$

If x is small, then the terms on the right are negligible and $e^{-x} \approx 1 - x$. With reference to the text, this approximation is $e^{-e^{\frac{-\alpha}{\beta}}} \approx 1 - e^{\frac{-\alpha}{\beta}}$.

[2] BOLENDER (2007) proposes that these anomalies reflect the fact that the Pirahã have lived in very small numbers (hundreds) since breaking away from the larger native population of the Mura by the early eighteenth century.

REFERENCES

ANDERSON, P. W. & ABRAHAMS, E. (2009): Machines fall short of revolutionary science. *Science,* 324, 1515–1516.

BEINHOCKER, E. (2006): *The Origin of Wealth: Evolution, Complexity, and the Radical Remaking of Economics.* Cambridge, MA: Harvard Business School Press.

BENTLEY, R. A. (2008): Random drift versus selection in academic vocabulary. *PLoS ONE,* 3(8), e3057.

BENTLEY, R. A., HAHN, M. W. & SHENNAN, S. J. (2004): Random drift and culture change. *Proceedings of the Royal Society B,* 271, 1443–1450.

BENTLEY, R. A., ORMEROD, P. & BATTY, M. (2010): Evolving social influence in large populations. *Behavioral Ecology and Sociobiology,* DOI: 10.1007/s00265-010-1102-1.

BENTLEY, R. A. & SHENNAN, S. J. (2003): Cultural evolution and stochastic network growth. *American Antiquity,* 68, 459–485.

BERGER, J. & LE MENS, G. (2009): How adoption speed affects the abandonment of cultural tastes. *Proceedings of the National Academy of Sciences,* 106, 8146–8150.

BOLENDER, J. (2007): Prehistoric cognition by description: A Russellian approach to the Upper Paleolithic. *Biology and Philosophy,* 22, 383–399.

BORGATTI, S. P., MEHRA, A., BRASS, D. J. & LABIANCA, G. (2009): Network analysis in the social sciences. *Science,* 323, 892–895.

BOYD, R. & RICHERSON, P. J. (1985): *Culture and the Evolutionary Process.* Chicago: University of Chicago Press.

BOYD, R. & RICHERSON, P. J. (1996): Why culture is common, but cultural evolution is rare. *Proceedings of the British Academy,* 88, 77–93.

BOYD, R. & RICHERSON, P. J. (2005): *The Origin and Evolution of Culture.* Oxford: Oxford University Press.

CHRISTAKIS, N. A. & FOWLER, J. H. (2009): *Connected: The Surprising Power of Our Social Networks and How They Shape Our Lives*. Boston: Little, Brown.

DUNBAR, R. I. M. (1993): Coevolution of neocortical size, group size and language in humans. *Behavioral and Brain Sciences,* 16, 681–735.

DURKHEIM, E. (1898): Représentations individuelles et représentations collectives. *Revue de Métaphysique et de Morale,* 6, 273–302.

EVERETT, D. L. (2005): Cultural constraints on grammar and cognition in Pirahã. *Current Anthropology,* 46, 621–646.

FRANK, S. A. (2009): The common patterns of nature. *Journal of Evolutionary Biology,* 22, 1563–1585.

GABORA, L. (2008): Mind. In: R. A. Bentley, H. D. G. Maschner and C. Chippindale (eds): *Handbook of Archaeological Theories*. Lanham, MD: AltaMira, pp. 283–296.

GALEF, B. G., JR. & LALAND, K. N. (2005): Social learning in animals: Empirical studies and theoretical models. *BioScience,* 55, 489–499.

GEARY, D. C. (2007): Educating the evolved mind: Conceptual foundations for an evolutionary educational psychology. In: J. S. Carlson and J. R. Levin (eds): *Educating the Evolved Mind: Conceptual Foundations for an Evolutionary Educational Psychology*. Charlotte, NC: Information Age, pp. 1–99.

GILLESPIE, J. H. (2004): *Population Genetics: A Concise Guide*. Baltimore: Johns Hopkins University Press.

GORDON, P. (2004): Numerical cognition without words: Evidence from Amazonia. *Science,* 306, 496–499.

GUIMERÀ, R., UZZI, B., SPIRO, J. & NUNES AMARAL, L. A. (2005): Team assembly mechanisms determine collaboration network structure and team performance. *Science,* 308, 697–702.

HAHN, M. W. & BENTLEY, R. A. (2003): Drift as a mechanism for cultural change: An example from baby names. *Proceedings of the Royal Society B,* 270, S1–S4.

HAMILTON, M. J. & BUCHANAN, B. (2009): The accumulation of stochastic copying errors causes drift in culturally transmitted technologies: Quantifying Clovis evolutionary dynamics. *Journal of Anthropological Archaeology,* 28, 55–69.

HAYEK, F. A. (1949): *Individualism and Economic Order*. London: Routledge.

HENRICH, J. (2004): Demography and cultural evolution: Why adaptive cultural processes produced maladaptive losses in Tasmania. *American Antiquity,* 69, 197–214.

HENRICH, J. (2006): Understanding cultural evolutionary models: A reply to Read's critique. *American Antiquity,* 71, 771–782.

HERZOG, H. A., BENTLEY, R. A. & HAHN, M. W. (2004): Random drift and large shifts in popularity of dog breeds. *Proceedings of the Royal Society B,* 271, S353–S356.

HRDY, S. B. (2009): *Mothers and Others*. Harvard University Press.

HULL D. L. (2001): *Science and Selection*. Cambridge: Cambridge University Press.

HYDE, J. S. & LINN, M. C. (2009): Gender similarities in mathematics and science. *Science,* 314, 599–600.

KANDLER, A. & LALAND, K. N. (2009): An investigation of the relationship between innovation and cultural diversity. *Theoretical Population Biology* 76, 59–67.

KING, R. D., ROWLAND, J., OLIVER, S. G., YOUNG, M., AUBREY, W., BYRNE, E., LIAKATA, M., MARKHAM, M., PIR, P., SOLDATOVA, L. N., SPARKES, A., WHELAN, K. E. & CLARE, A. (2009): The automation of science. *Science,* 324, 85–89.

KLEIN, R. (2002): *The Dawn of Human Culture*. New York: Wiley.

LALAND, K. N. (2004): Social learning strategies. *Learning Behavior,* 32, 4–14.

LALAND, K. N. & READER, S. M. (2010): Comparative perspectives on human innovation. In: M. J. O'Brien and S. J. Shennan (eds): *Innovation in Cultural Systems: Contributions from Evolutionary Anthropology*. Cambridge, MA: MIT Press pp. 37–51.

LEVINSON, S. C. (2005): Comment on 'Cultural constraints on grammar and cognition in Pirahã: Another look at the design features of human language,' by D. L. Everett. *Current Anthropology,* 46, 637–638.

MESOUDI, A. & LYCETT, S. J. (2009): Random copying, frequency-dependent copying and culture change. *Evolution and Human Behavior,* 30, 41–48.

MITHEN, S. J. (1996): *The Prehistory of the Mind: The Cognitive Origins of Art, Religion, and Science*. London: Thames and Hudson.

NEIMAN F. D. (1995): Stylistic variation in evolutionary perspective: Inferences from decorative diversity and interassemblage distance in Illinois Woodland ceramic assemblages. *American Antiquity,* 60, 7–36.

O'BRIEN, M. J., LYMAN, R. L. & SCHIFFER, M. B. (2005): *Archaeology as a Process: Processualism and Its Progeny.* Salt Lake City: University of Utah Press.

POWELL, A., SHENNAN, S. & THOMAS, M. G. (2009): Late Pleistocene demography and the appearance of modern human behavior. *Science,* 324, 1298–1301.

RENFREW, C. & SCARRE, C. (eds.) (1998): *Cognition and Material Culture: The Archaeology of Symbolic Storage*. Cambridge: McDonald Institute for Archaeological Research.

RICHERSON, P. J. & BOYD, R. (2005): *Not by Genes Alone*. Chicago: Chicago University Press.

RICHERSON, P. J., BOYD, R. & BETTINGER, R. L. (2009): Cultural innovations and demographic change. *Human Biology,* 81, 211–236.

RIDLEY, M. (2010) *The Rational Optimist: How Prosperity Evolves*. London: Fourth Estate.

SALGANIK, M. J., DODDS, P. S. & WATTS, D. J. (2006): Experimental study of inequality and unpredictability in an artificial cultural market. *Science,* 311, 854–856.

SHENNAN, S. (2000): Population, culture history, and the dynamics of change. *Current Anthropology,* 41, 811–835.

TOMASELLO, M., KRUGER, A. & RATNER, H. (1993): Cultural learning. *Behavioral and Brain Sciences,* 16, 495–552.

TURCOTTE, D. L. (1997): *Fractals and Chaos in Geology and Geophysics* (sec. ed.). New York: Cambridge University Press.

ZILHÃO, J. ANGELUCCI, D. E., BADAL-GARCÍA, E., D'ERRICO, F., DANIEL, F., DAYET, L., DOUKA, K., HIGHAM, T. F. G., MARTÍNEZ-SÁNCHEZ, M. J., MONTES-BERNÁRDEZ, R., MURCIA-MASCARÓS, S., PÉREZ-SIRVENT, C., ROLDÁN-GARCÍA, C., VANHAEREN, M., VILLAVERDE, V., WOOD, R. & ZAPATA, J. (2010): Symbolic use of marine shells and mineral pigments by Iberian Neandertals. *Proceedings of the National Academy of Sciences,* 107, 1023–1028.