9-1-2015

Cultural Incubators and Spread of Innovation

Enrico R. Crema
UCL Institute of Archaeology, enrico.crema@upf.edu

Mark W. Lake
UCL Institute of Archaeology, mark.lake@ucl.ac.uk

Cultural Incubators and Spread of Innovation

Enrico R. Crema[1,2,3] and Mark W. Lake[1]*

[1]UCL Institute of Archaeology.

[2]Universitat Pompeu Fabra, Department of Humanities.

[3]CaSEs - Complexity and Socio-Ecological Dynamics Research Group.

*Correspondence to: Mark W. Lake UCL Institute of Archaeology 31-34 Gordon Square London WC1H 0PY UK. E-mail: mark.lake@ucl.ac.uk.</FN>

Short Title: Cultural Incubators and Spread of Innovation

KEY WORDS: CULTURAL TRANSMISSION, PAYOFF-BIASED TRANSMISSION, PAYOFF UNCERTAINTY, SPREAD OF INNOVATION, AGENT-BASED MODEL.

Abstract Several forms of social learning rely on the direct or indirect evaluation of the fitness of cultural traits. Here we argue, via a simple agent-based model, that payoff uncertainty, that is the correlation between a trait and the signal used to evaluate its fitness, plays a pivotal role in the spread of beneficial innovation. More specifically, we examine how this correlation affects the evolutionary dynamics of different forms of social learning and how each can generate divergent historical trajectories depending on the size of the sample pool. In particular, we demonstrate that social learning by copying the best model is particularly susceptible to a sampling effect caused by the interaction of payoff uncertainty, the number of models sampled (the sample pool) and the frequency with which a trait is present in the population. As a result, we identify circumstances in which smaller sample pools can act as 'cultural incubators' that promote the spread of innovations, while more widespread sampling of the population actually retards the rate of cultural evolution.

Social learning is the key process underlying the spread of cultural traits in a living population. The last thirty years have seen productive application of models inspired by population biology to specific questions of interest to anthropologists, archaeologists and ethologists, including: what evolutionary 'problem' social learning solves (Boyd and Richerson 1988, 1995; Perreault 2012; Plotkin and Odling-Smee 1982); the spread of innovations (Henrich 2001; Kandler and Steele 2009); the loss of cultural traits (Henrich 2004); the effect of population size on cultural complexity (Shennan 2001; Powell et al. 2009, 2010); and the rationality of human decision-making (Bentley and Ormerod 2012). The latter, in particular, points to a convergence of interest with economists interested in bounded rationality which, although more commonly investigated using the mathematical apparatus of game-theory, exhibits population thinking (Boyd and Richerson 2000) in

the explicit use of replicator dynamics to explore optimal strategies for copying with payoff uncertainty (Schlag 1998). In this paper we cross-cut some of these established themes by exploring the impact of population size/sampling and payoff uncertainty on the early stages in the spread of innovations. Our finding augments the results of some existing studies (e.g., Baldini 2012, 2013; Lake and Crema 2012), but also identifies circumstances in which the positive correlation between population size and the rate of cultural evolution (Henrich 2004; Powell et al. 2009, 2010; Shennan 2001) is significantly modified by the joint effect of payoff uncertainty and sample pool size. We begin with a brief review of findings from the most immediately relevant recent literature on the value of social learning, the nature of payoff and success-biased social learning and how it is mediated by payoff uncertainty and population size and/or the way in which the population is sampled.

The Value of Social Learning. The textbook definition of social learning is copying some cultural trait from another individual rather than independently developing it by trial-and-error. The extent to which copying from others is unambiguously separable from trial-and-error—also called individual or 'asocial' (Laland 2004)—learning is debated in terms of the underlying psychological mechanisms (Heyes and Plotkin 1989) and may vary by context (Mesoudi et al. 2014) and, indeed, the window of observation, since an individual can potentially copy a trait, modify it and then pass it on to another individual—so-called 'guided variation' (Boyd and Richerson 1985). Furthermore, it is now well understood that individual learning and social learning are interdependent at population level. Evolutionary epistemologists had long surmised that the function of social learning is to bring about adaptive responses to selective pressures over a timescale intermediate between (slow) natural selection and (fast) individual learning (Campbell 1974; Odling-Smee 1983,

but see Kendal et al. 2009 for a cautionary note), but it was Boyd and Richerson's (1985, 1988) 'costly information hypothesis' which identified the selective benefit to individual organisms of social learning: that it incurs a lower cost than individual learning, provided that the rate of spatial and temporal variation in the environment is not so great as to render the socially learned information useless. Rogers (1988) subsequently demonstrated, however, that social learning does not increase the average fitness of the population as a whole, because imitators are selected against once their lower cost of learning is offset by the benefit of more accurate environmental tracking, in other words, when the fitness of the two types is equal. Further analysis (Boyd and Richerson 1995) upheld Rogers' conclusions, with the result that research is now focussed on the existence of mixed equilibrium distributions of both individual and social learners (e.g., Kendal et al. 2009; Rendell et al. 2010), and on the circumstances under which individuals should engage in social learning (Laland 2004; Kandler and Laland 2013; Kendal et al. 2005; Kendal et al. 2009).

Social Learning and Payoff Signals. Even if beneficial social learning requires at least some individual learning, it is nevertheless clear that many behaviours are predominantly acquired by copying (Bandura 1977; Boyd and Richerson 1985). Laboratory experiments have revealed the extent to which human beings are cognitively predisposed to copy from others (Herrmann et al. 2007) and the extent to which cultural norms reflect that predisposition has been demonstrated by comparative field studies (e.g., Rice and Feldman 1997; Hewlett et al. 2002). Furthermore, Henrich (2001) demonstrated that biased social learning that excludes individual trial-and-error is required to account for the 'S-shape' curves that describe the uptake of the vast majority of well-studied innovations. There has been significant interest lately in the explanatory power of models based on random social copying (Bentley and Shennan 2005; Bentley and Ormerod 2012) and we include

this in a variant of our model, but our principal interest is forms of social learning in which the selection of a cultural trait from a set of alternatives in a sample pool is biased by direct or indirect (Boyd and Richerson 1985) estimation of its 'worth'.

Contemporary social learning nomenclature distinguishes 'content' bias from 'context' bias (Henrich and McElreath 2003). Content biases are those which arise from the intrinsic properties of a trait, either because learners have a psychological predisposition (either biologically or culturally determined) toward certain kinds of information, or because they engage in "more generalized cost-benefit calculation" (Henrich and McElreath 2003: 129). Recently, Mesoudi (2011: 67) has suggested that all 5 of the characteristics of successful innovations identified by Rogers (2003) can be considered forms of content-bias. Traits are more likely to be adopted if they are: (i) advantageous relative to existing traits; (ii) compatible with existing traits; (iii) simple enough that their use is easily learned; (iv) testable, so that their advantage can be discerned; and (v) readily observable. Henrich and McElreath's notion of "cost-benefit calculation" and Rogers' "relative advantage" are both examples of what is elsewhere termed 'payoff bias' (Kendal et al. 2009) and is the form of social learning most commonly studied by economists (e.g., Schlag 1998).

Context biases arise when learners utilise cues in the learning environment as proxies for the 'worth' of a trait: either its frequency or the characteristics of those individuals (models) who already possess that trait. Frequency dependent bias occurs when the probability of an individual copying the most common trait is higher (conformism) or lower (anti-conformism) than it would be if they randomly sampled the population of traits. Although frequency dependent social learning is in itself a low-cost form of learning, it is particularly susceptible to the trade-off, already discussed above, between low learning cost and the potentially high cost of failure to adequately track the

environment (Wakano and Aoki 2007; Kandler and Laland 2013). The other form of context bias is model-biased social learning, which entails treating the prestige (Henrich and Gil-White 2001), success (Baldini 2012; Henrich and McElreath 2003), similarity, or some other property of a potential cultural model as a proxy for the 'worth' of a cultural trait which that model possesses.

Although success bias is classified as a context bias because the learner evaluates a trait indirectly via its contribution to the reproductive fitness or material well-being of individuals who possess it (Baldini 2012; Henrich and McElreath 2003) there might be circumstances in which the result would not be very different from direct evaluation of the trait payoff, which is classified as a content bias. Indeed, the structural similarity of the two forms of bias is attested by the fact that the same game-theoretic analysis by Schlag (1998) is cited in both Henrich and McElreath's (2003) discussion of success bias and Kendal et al's (2009) discussion of payoff bias; in fact, there are no terms in Schlag's model which would unambiguously differentiate the two forms of bias. Perhaps the best way of conceiving what is at stake in the attempt to differentiate payoff bias, success bias and prestige bias is to think in terms of a spectrum in which evaluation of the trait is increasingly prone to false inference about its true worth. Put another way, while all three biases entail paying attention to 'payoff signals', they differ in the likely correlation between the signal and the target trait owing to an increasing number of factors that may contribute to the generation of the payoff signal. Thus payoff biased social learning is vulnerable to stochasticity inherent in the use or performance of a trait, while success bias layers on top of this the possibility that the model is systematically better or worse equipped than the learner to use the trait. For example, a hunter might measure the advantage of a given target trait (e.g., the shape of an arrowhead) via the success (e.g., number of prey captured) of its bearer (cf. Mesoudi and O'Brien 2008a, 2008b). This will, however,

depend on the type of bow used, the strength and skill of the hunter and their knowledge of prey

species, layered on top of stochastic factors that might equally affect both learner and model, such

as the availability of the prey and sheer luck. All these factors will contribute to the payoff signal,

so that two hunters with identical arrowheads will capture different numbers of prey. Prestige bias

renders evaluation of the trait even more indirect because learners defer evaluation of the success of

the model to others (Henrich and Gil-White 2001, Atkisson et al. 2012). For this reason, in our

model we do not explicitly differentiate between payoff bias and success bias, preferring instead to

focus on the magnitude of what, following Baldini (2012), Schlag (1998) and others, we simply

label 'payoff uncertainty'.

The Effect of Payoff Uncertainty   Payoff uncertainty arises when the payoff signal is

noisy, which may be because of inherent stochasticity and/or because the signal is also a function of

the context in which it was generated, for example other cultural and biological traits possessed by

the model, or the environment where the trait is manifest. A number of studies have modelled

payoff uncertainty by assuming that a given cultural trait g generates a payoff signal p drawn from a

parametric probability distribution $f(p)$, such that the variance can be regarded as a measure of the

correlation between the signal and the target trait.

Boyd and Richerson (1995) adopted a variant of this approach in a model designed to

explore whether copying can increase the average fitness of learners by allowing individuals to

learn more selectively in a changing environment. Boyd and Richerson modelled payoff uncertainty

in the result of individual—rather than social—learning, such that the observed difference between

the payoff of each of two traits is a normal random variable. In these circumstances learners can

ensure greater accuracy by insisting on a greater difference between the observed payoffs before

adopting the trait with the higher apparent payoff, but this will also result in more cases where the difference threshold is not met and so the result of their individual learning is indeterminate, in which case they copy a randomly chosen individual. What Boyd and Richerson found was that the average fitness of a population at equilibrium in terms of the frequency of copying could be higher than that of a population of learners who do not copy, but only for particular probabilities of environmental change.

Boyd and Richerson's model employed unbiased social learning, but more recently Kendal et al. (2009) described a model in which increasing the payoff associated with the novel trait allowed payoff-biased social learning to evolve in environments characterised by a higher probability of change. Although, as Baldini (2012) notes, Kendal et al's model of payoff is not stochastic, there is nevertheless a sense in which the relative payoff of the two traits is at least a partial measure of the strength of the payoff signal. Baldini's (2012) own model of the evolution of success-bias does employ stochastic payoffs, such that the successes associated with the two traits are independently and normally distributed. Baldini demonstrated that if the measure of success used by social learners is also that upon which natural selection acts to alter the frequency of social learning strategies, then increasing the strength of selection favours greater reliance on success bias since it effectively also increases the strength of the payoff signal by better differentiating the average payoffs of the competing traits.

Baldini's model of success-bias assumes that the learner pays attention to the mean payoff of all the sampled models possessing each unique cultural trait, although the extent to which the learner actually favours higher mean payoffs is parametrised. This is a natural choice for a model extended to explore the evolution of success bias when the payoff signal is also that upon which

natural selection acts, but as he concedes, it is not the only possible success-biased learning rule. Other possibilities include copying the most successful model (copy-the-best), copying a model who is more successful than the learner (copy-if-better) and copying with a probability that is proportional to how much more successful the model is relative to the learner (copy-proportionally) (Laland 2004).

The copy-the-best and copy-if-better rules both suffer from an inability to differentiate between "lucky and certain (or highly probable) payoffs" (Schlag 1998: 142). Indeed, Baldini (2013) has recently shown how a social learning strategy based on copying the best individual (Schlag 1999; Apesteguia et al. 2007) can occasionally lead to the spread of suboptimal traits if trait payoff distributions overlap, but the variance of the payoff is higher for the suboptimal variant. Furthermore, Lake and Crema (2012) have demonstrated that even if the variances are equal, payoff uncertainty can result in copy-the-best learning failing to identify a superior trait if it is rare, simply because the rarer the superior trait is, the more likely that at least one model possessing the inferior trait will have obtained a higher payoff. It turns out that this sampling effect reduces the optimal amount of payoff-biased copy-the-best social learning when there is payoff uncertainty because it causes the early loss of innovations, although Lake and Crema go on to suggest (but did not formally explore) that this effect is also a function of the fraction of the total population of potential models that is actually sampled (the size of the 'sample pool').

Copy-proportionally serves to mitigate false inferences arising from the 'lucky' realisation of improbable payoffs (Schlag 1998), as does copying the best *average payoff* of each cultural variant (as per Baldini 2012; see also Schlag 1999). The latter 'copy-the-best-average' algorithm is not affected by the difference in payoff variance so long as the sampling is truly random such that

models with higher payoff are no more likely to be used to evaluate the average payoff of each variant than other models. Although success-biased rules by which learners sample multiple models and then average the payoffs associated with competing traits have the advantage of suppressing false inference from improbable payoffs, Baldini (2012) has demonstrated that—perhaps counter-intuitively—they are also biased towards rare traits, with the effect that they can perform less well than unbiased learning when the optimal trait is common. Baldini observes that the bias towards rare traits is a consequence of the fact that by considering the average payoff of each variant, the choice between them is effectively made on the basis of only one observation per variant. Thus, for example, if trait A is twice as frequent as trait B, but both have the same payoff distributions, then the probability of choosing between them on the basis of their average payoffs is nevertheless close to 0.5; consequently, the rarer trait B will be chosen with greater probability than its frequency, so altering the trait distribution in favour of the rarer trait.

The Effect of Population Size.　　　　The effect of population size on cultural evolution resulting from payoff/success-bias has been formally modelled in several studies (Henrich 2004; Powell et al. 2009, 2010; Shennan 2001), with the general conclusion that larger populations promote cultural evolution by accelerating adaptive change and/or permitting the spread of more complex (harder to learn) cultural traits. A recent social learning experiment by Derex et al. (2013) claims to provide empirical support for the positive effect on cultural evolution of increasing group size, although Andersson and Read (2014) have disputed this, pointing out that the experimental result that larger groups produced complex artefacts more frequently than smaller groups could simply be a sampling effect.

　　　　Shennan (2001) presented two models of the innovation and spread of fitness enhancing

craft skills in different size populations. In both cases he modelled the fitness of an individual as the multiplicative sum of the states of each of a fixed number of craft attributes possessed by that individual. The states of these individual craft attributes could be altered by innovation, such that most innovations have relatively little effect but some have a significant effect (ibid. p6-7). In the first model craft attributes are passed vertically from parent to offspring, so that change in the frequency of any given attribute state is a function of the differential reproduction of the individuals who possess it and whose fitness is partially dependent upon it. The second model added oblique cultural transmission such that individuals probabilistically either acquire attribute states from their same sex parent or from a pool of k randomly chosen cultural models. If $k = 1$ social learning is effectively unbiased transmission, but if $k > 1$ the individual copies the most attractive individual among the k models, thus introducing a copy-the-best learning rule, albeit one that does not appear to include any evaluation of whether the model is actually better than the individual who is copying. This cultural transmission process occurs on a per-attribute basis, such that the most attractive individual among the k models is chosen with respect to the trait in question (ibid. p10). Using what he considered realistic parameters for hunter-gatherers, Shennan found that increasing the population size increased the long-term geometric mean fitness of the population, with the effect most pronounced for smaller sizes and less so for larger population sizes; this was true for both models. The effect of oblique transmission was to increase the absolute geometric mean fitness for any given population size (at least at the innovation rate for which explicit comparison is provided), but the positive effect of increasing population size decreases as the rate of oblique transmission is increased (ibid. fig 3).

Shennan's broad conclusion about the positive correlation between population size and

cultural evolution was supported by Henrich's (2004) model of the social learning of skills. Henrich modelled oblique transmission of skills in which individuals attempt to copy a skill from the most skilful individual in the population, in other words, using a copy-the-best social learning rule. He also analysed a variant which incorporated a fraction of vertical transmission, but this produced the same qualitative results. Importantly, and central to the claim that his model is "rooted in the cognitive details of social learning and inference" (ibid: 197), Henrich's model uses the Price equation, thereby explicitly incorporating the effects of both cultural selection and errors in social learning. In particular, he modelled the latter by drawing the skill-level achieved by the learner from a Gumbel distribution, such that the copied skill level is most likely to be lower than the model's skill, but that there is nevertheless some non-zero probability of it being higher; in other words, Henrich assumed that social learning is mostly imperfect, usually with deleterious consequences, but occasionally producing improvements. By varying the mode, α, and dispersion, β, of the Gumbel distribution, Henrich was able to separately parameterise the difficulty of learning a particular skill (its complexity) and the likelihood that those attempting to copy it make similar or potentially very different errors (in a sense ease or difficulty of innovating). Henrich analysed the size of the population required for the change in average skill to be positive, that is, for there to be "cumulative adaptive cultural evolution" (ibid: 202), and found that it depends on the ratio between the difficulty of learning and ease of innovating. Skills that are more complex (large α) and/or more difficult to innovate (low β) are only maintained or cumulatively improved in larger populations.

Shennan and Heinrich's seminal contributions to understanding the importance of population size for cultural evolution have since been extended by Powell et al. (2009, 2010). Powell et al. were motivated by the same anthropological problem that initially motivated

Shennan's 2001 model, the so-called 'explosion' of cultural complexity at the European Middle-Upper Palaeolithic Transition. In order to explore the plausibility of demographic change as a prime driver of this transition, Powell et al. (2009) re-implemented Henrich's 2004 model, albeit with three changes. First, and of least significance for our purposes, they simplified the treatment of errors in social learning by collapsing α and β into one parameter. Second, they eschewed Henrich's simple oblique transmission model with one which they claim better captures a more realistic assumption that "cultural skills and behaviours are often first, and sometimes only, learned from parents" and also that "identification of the most skilled model will be inaccurate" (Powell et al. 2009: supplementary material). This entails "vertical" unbiased copying of skills from individuals in the previous generation followed by 'oblique" biased copying from individuals in the same generation (ibid.). The bias in the 'oblique' copying results from a social learning rule in which each individual attempts to copy a skill from among those models who have a higher skill-level than that which it inherited from the previous generation. The chosen model is selected probabilistically in proportion to the magnitude of the difference its skill-level and that of the learner. Third, Powell et al. added spatial structure to the population by dividing it into G groups, each of size N, placed in an environment. Individuals can migrate between groups at a rate which is a function of the density, D, with which the groups are placed in the environment. Powell et al. found that increasing the overall population size (a function of G, since N is fixed) does promote increased skill complexity, but the effect is most pronounced for smaller population sizes and is much diminished or even non-existent for larger population sizes (ibid:fig. S2). In the latter case (G > 50) the principal determinant of population average skill-level is the amount of migratory activity. Experiments with a fixed and large number of groups demonstrated that the positive effect of increased inter-group

migration on skill-level is greater for more complex skills, in the sense that as skills become harder to learn so increasing the migration rate has an effect that is the equivalent of greatly increasing the size of a single unstructured population. Although Vaesen (2012) has argued that the assumptions made by Powell et al. mean that their model does not provide a sufficiently severe test of their substantive hypothesis, their results do nevertheless extend the set of specific learning rules for which it appears that there is a positive correlation between increasing population size and/or density (see Powell et al. 2009, fig. S1) and the rate at which payoff/success-biased social learning can drive cultural evolution. Furthermore, their study also points to the importance of population structure for this relationship.

Combining Payoff Uncertainty and Sampling Effects. We suggested above that payoff uncertainty potentially affects any form of social learning that makes use of a payoff-signal to evaluate the worth of a cultural trait, whether that is content-biased social learning such as payoff-biased learning, or context-biased social learning such as success or prestige-biased learning. In all these cases a simple stochastic model of payoff can be used to capture the fundamental structural issue, which is the strength of the correlation between the observed payoff signal and the true 'worth' of the trait. As we have discussed, several studies employing such a model have demonstrated that payoff bias significantly affects the optimal amount of social learning in a range of different scenarios. Moreover, it is also now clear that the exact form of the social learning rule matters: copy-the-best can lead to false inferences when payoff uncertainty is high, whereas copy-the-best-average payoff is more robust, but on the other hand favours rare variants.

We have also documented how a separate strand of theoretical modelling has demonstrated that population size and/or the way in which the population is sampled mediates the effect of

payoff/success-biased learning, and this appears to be true for both copy-the-best and proportional copying rules. Deleterious sampling effects caused by small population sizes are predicted to reduce the rate of adaptive cultural evolution and even lead to cultural trait loss. Larger populations, on the other hand, are predicted to support the social-learning of more complex traits even when the learning process is error prone, and increasing the migration rate in populations broken into subgroups also has this effect. It has recently been noted, however, that empirical studies do not unambiguously support these theoretical insights (Collard et al. 2013). There are various possible reasons for this, ranging from the effect of population size being swamped by other factors, to the empirical studies not adequately measuring the interacting as opposed to total (census) population. The possibility that particularly interests us here is that existing models of the effect of population size/sampling on cultural evolution do not adequately take account of the effect of payoff uncertainty on social learning. This is particularly problematic if Caldwell et al. (2014) are correct in their supposition that social learning of human technology invariably involves payoff uncertainty because of the spatio-temporal separation of tool manufacture from tool use.

While it is clear from existing studies that both payoff uncertainty and population size/sampling alter the effect of social learning on cultural evolution, it is striking that all the models reviewed above incorporate one or the other, but not both, so their interaction has not been fully explored[1]. Furthermore, in our discussion above we noted studies (Baldini 2012; Lake and Crema 2012) which provide reason to think that when payoff is uncertain the efficacy of particular forms of social learning can also be affected by the relative frequency with which a 'better' trait is

---

[1] Appendix D of Henrich (2004) presents a variant of his model in which selection of the cultural model is based on an indirect cue for the underlying skill, but this is not explored in the body of the paper and, in any case, he was not concerned with population structure.

present in a population. In particular, we have previously demonstrated that a rare better trait can be lost as a result of payoff uncertainty because of a sampling effect in which if enough models with the more common inferior trait are sampled it becomes probable that one of them will have a rare high payoff that is greater than the highest payoff exhibited among the smaller sample of models with the rare but better trait (Lake and Crema 2012). This effect suggests to us a mechanism which could reduce the benefit of larger population sizes, or of higher migration rates between subgroups, especially when the better trait is rare, as would be the case in the earliest stages of the diffusion of an innovation. In section 2 of this paper we describe a model designed to explore exactly that possibility and in section 3 we present results which confirm that there are indeed scenarios in which certain payoff/success-biased social learning rules are predicted to produce faster rates of cultural evolution in smaller rather than larger populations. Moreover, it seems that smaller communities within populations can act as initial 'cultural incubators' which promote the early survival of beneficial innovations, and we note the far-reaching significance of this in our concluding comments in Discussion and Conclusion.

Modelling Cultural Transmission with Uncertain Payoff Signals

We first describe a model of the cultural transmission of discrete traits whose payoff is variable, such that individuals have imperfect knowledge of the true worth of traits possessed by others in the population. Three variants of this model differ according to the learning rule employed, which is one of: random copying (unbiased transmission), copy-if-better and copy-the-best. We provide a commentary on aspects of the model and then go on to outline our experimental design.

The Model. Consider a population of N individuals, each possessing a cultural trait g represented by an integer, and a payoff signal p that varies at each time-step and is a function of g.

We assume that neither g nor p affects the reproductive fitness of the agents (i.e. N is held constant), and that the simulation proceeds synchronically rather than sequentially (i.e. all agents execute each step before proceeding to the subsequent one). At each discrete time-step, t, all individuals undertake the following:

1. *Expression of the payoff-signal*: a payoff signal p_i is emitted by each individual as a random draw from a normal distribution with mean g_i and standard deviation σ. We assume that σ is equivalent for all values of g.

2. *Social learning*: With some probability z, each individual, i, engages in social learning which may result in it copying a trait from a model chosen from among k individuals who comprise the sample pool, S_i. The choice of model and decision whether to adopt the model's trait is governed by one of three learning rules (all individuals use the same learning rule in any given experiment)

   a. *Random copying* (RC). The focal individual, i, randomly selects a model, j, from the sample pool and adopts that model's trait irrespective of its properties or payoff.

   b. *Copy-if-better* (CIB). The focal individual, i, randomly selects a model, j, from the sample pool. It then compares its own payoff, p_i with the model's payoff, p_j, and adopts the model's trait only if $p_j > p_i$ (otherwise it retains its original trait).

   c. *Copy-the-best* (CB). The focal individual selects the model, m, with the highest payoff, p_m, among the k individuals in the sample pool S_i. The focal individual subsequently compares its own payoff, p_i, with that of the model and if $p_i < p_m$, it

17

copies the cultural trait g_m of the model.

3. *Innovation.* For the sake of simplicity we assume innovation to be incremental and beneficial, the latter solely in the sense that it entails a higher cultural selective advantage. Innovation occurs with frequency μ and consists of an update of cultural trait g_i by an amount b so that at time t, $g_i(t) = g_i(t\text{-}1) + b$. Note that we allow for convergence, so two agents can, with probability μ^2, independently reach the same $g(t)$ if they also had the same $g(t\text{-}1)$.

As discussed in section 1, the stochasticity of the payoff signal (σ) is a direct proxy for the correlation between g and p. When σ=0, there is no uncertainty in the payoff, and agents are assumed to be capable of fully isolating the adaptive performance of a given cultural trait, and hence can correctly identify whether a variant is "better" or the "best," depending on the social learning rule. With increasing values of σ, the correlation declines and consequently traits with lower g can occasionally be associated with a payoff signal that is higher than that associated with other traits that have higher g.

Random copying (RC; unbiased transmission) provides a benchmark expectation for a comparative assessment of the two social learning rules that incorporate selection. Random copying assumes a copying process in which only the frequencies of the variants determine their probability of adoption (Bentley et al. 2004). The *copy-the-best* rule (CB) involves a process of selection (the identification of the individual in the sample pool who has the highest payoff, p) and evaluation (comparison of the focal individual and model's payoff signals) prior to copying. Copy-if-better (CIB) offers a compromise between undirected random copying and the strongly directed copy-the-

best rule. On the one hand there is no directed selection of the model (which is picked at random from the sample pool), but on the other hand copy-if-better retains the element of evaluation in that the model's trait is adopted only if it has a higher payoff, p, than the focal individual's current trait. We noted in section 1 that Powell et al. (2009), doubting the realism of the copy-the-best rule, implemented a form of copy-if-better social learning in which individuals have an opportunity to copy a better skill with a probability proportional to the magnitude of improvement it represents. It could be argued that for Powell et al. the probabilistic copying provides a partial implementation of payoff uncertainty—in the sense that higher skill-levels are more likely to be perceived as worth copying—, but on the other hand, it does not ultimately capture the uncertainty about which models have higher skills. Our version of copy-if-better allows payoff uncertainty to dictate the accuracy of inference about which models have higher skills and is more weakly directed than Powell et als' rule, but on the other hand it should be less susceptible to false inference from rare payoffs than both their rule and the copy-the-best rule.

An important element in our model is the parameter k, the number of individuals in the sample pool. This allows us to relax the assumption that social learning is unconstrained by physical or cognitive constraints (i.e. that cultural transmission is panmictic) since it implies that the sampling capacity of the focal individual is finite, perhaps as a function of cognitive capacity (cf. Dunbar 1993), or the result of cultural isolation caused by distance (cf. Premo and Scholnick 2011). Both Shennan (2001, second model only) and Lake and Crema (2012) parameterised the size of the sample pool, but neither subjected it to systematic exploration and Shennan was not, in any case, concerned with the effect of payoff uncertainty. Powell et als' (2009) use of discrete groups (subpopulations) connected by migration is analogous to specifying finite sample pools with non-

identical but partially overlapping membership, but again, they were not explicitly concerned with the relationship between sample pool size and payoff uncertainty. The parameter k can be loosely interpreted as the size of the strongly interacting social clique who learn from one-another, which potentially varies from the entire population of size N, to diads comprising just 2 individuals; however, in our formalisation we always exclude the focal individual (the one who is learning) from the sample pool, so k can take values from N-1 to 1. In this way our model assumes that the focal individual never attempts to copy from itself, although it can subsequently evaluate a candidate trait against its own trait (in copy-the-best and copy-if-better social learning). Consequently, model selection and evaluation (the decision to adopt a candidate trait) are explicitly treated as separate parts of the social learning process.

Experimental Design. In this paper we present the results of three experiments. The first explores how different values of σ (uncertainty in the payoff signal), k (the size of the sample pool) and the social learning rule interact to affect the earliest stage of beneficial innovation. Our specific concern is to establish the probability that a rare beneficial innovation is lost from the total population, N, as a result of sampling effects. We achieve this by assuming that there is one innovator who has a superior trait and that all other individuals in the population share an inferior trait (with lower g and therefore lower mean p). For simplicity we assume that the innovation and the existing trait both have the same payoff variance, σ. In this first experiment we are not concerned with population sampling, so in all cases $k = N$ - 1. In other words, the sample pool is simply the entire population minus the focal individual, and thus by altering k we are effectively altering the population size. In this way, this first experiment shares with Shennan (2001) and Henrich (2004) a principal interest in

the effect of overall population size[2] rather than different sampling fractions.

The second and third experiments focus on how different values of σ (uncertainty in the payoff signal), k (the size of the sample pool) and the social learning rule interact to determine the long-term rate of cultural evolution. The latter is measured as the average value of g after a pre-defined number of time-steps. In other words, we seek to identify conditions where, holding constant the rates of innovation (μ) and the amount of improvement (b), we see the fastest spread of beneficial cultural traits. Furthermore, in this case we abandon the assumption of panmixia and instead explore the effect of population sampling. We achieve this by implementing the model as a simple agent-based simulation in which we randomly place the agents in a bounded space and define the sample pool of each focal agent as its k nearest-neighbours (see figure 1). Thus, when $k < N - 1$ we model some degree of isolation by distance whereas when $k = N - 1$ we model a panmictic population. Note that modelling the membership of the sample pool as a fixed number of neighbours (rather than, say, those falling within some euclidean search radius) guarantees a fixed size sample pool per agent and, importantly for interpretation of N, that there are no completely isolated population subgroups. In experiment 2 we hold the total population, N, constant and vary k, hence we are no longer modelling the effect of population size *per se*, but rather the completeness with which the population is sampled. In that regard this second experiment is closer to some of the experiments conducted by Powell et al. (2009) than it is to the earlier studies of Shennan (2001) and Henrich (2004), although of course it is different in its very explicit treatment of payoff uncertainty.

---

2   Although if one is concerned with absolute numbers it should be noted that both Shennan and Henrich report results for N/4, reflecting their shared assumption that individuals predominantly learn from same-sex parents (Henrich adds a further 20% reduction to reflect the reduced likelihood of learning from the oldest potential models). This is really just scaling for the purposes of anthropological comparison; indeed, although both Shennan and Henrich refer to N/4 as the "effective population" they do not appear to mean that in the technical sense of it being equal to the size of a Fisher-Wright population that would exhibit the same amount of drift.

In experiment 3 we vary both the census population size (*N*) *and* the sample pool size *k*. Table 1 records the range of parameter values used in the three experiments.

[Table 1]

Table 1: Model parameters and values used in the three experiments.

Symbols	Description	Exp. 1	Exp.2	Exp. 3
N	Number of agents	2-151	1,000	1,000, 2000, and 4000
σ	Uncertainty of the payoff signal	0.5, 1	0-3	0, 0.4, 1.5, and 4
z	Frequency of social learning	1	1	1
k	Size of the sample pool of potential model agents	*N*-1	1-500	1-500
μ	Innovation rate	-	0.005	0.005
c	Convergent innovation rate	0, 0.005	-	-
b	Amount of beneficial innovation	1	1	1

[Figure 1]
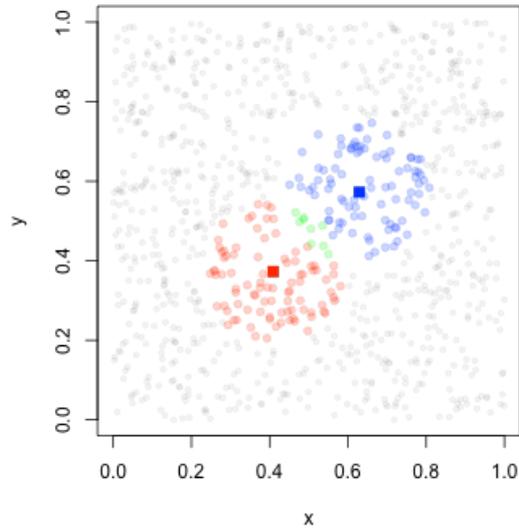
Figure 1: Schematic illustration of how *k* is defined in experiments 2 and 3. Focal individuals are depicted as solid squares, with their respective pool of social teachers (the *k* nearest neighbours) shown as round dots of the same colour. Green dots represent the shared portion of the pool, while the grey dots are individuals which fall outside the sampling pool (*N*=1000, *k*=100).

Results

Experiment 1: Loss of a Beneficial Innovation. When the mode of social learning is random copying we can calculate the probability of the population losing the innovation, P(*loss of B*), as follows. First, recall that we assume that one individual has innovated a better trait, B, and all other $N - 1$ individuals possess an inferior trait, A. Recall also that we fix $k = N - 1$, which means that individuals do not copy themselves. The probability that the population loses B is the product of the probability, P(*innovator loses B*), that the innovator loses B by copying A and the probability,

P(*others do not copy B*), that none of the other N - 1 individuals copies B from the innovator. Considering each in turn:

P(*innovator loses B*) = z, because the innovator will always loose B if it engages in unbiased social learning as all other individuals from which it can learn have trait A;

P(*others do not copy B*) = $(1-z/k)^k$, because the probability of any one of the $k = N$ - 1 individuals who are not the innovator not copying trait B is $1-z/k$.

It follows that:

P(*loss of B*) = $z(1-z/k)^k$

This solution does not, however, allow for the possibility that each of the individuals who has trait A might gain B, not by copying it from the innovator, but by convergent innovation of its own. Although this possibility might be remote (but perhaps less so for simpler traits), we take a conservative stance by incorporating it: if the innovation is lost even with the possibility of convergence then the probability of loosing it without convergence would be even higher. Assuming that copying trait B and independently innovating it are not mutually exclusive possibilities and that the probability of convergent innovation is c, then by the law of addition, the probability of any one individual who is not the original innovator acquiring the trait B is $z/k+c$-cz/k.

Substituting this revised formulation back into our original solution, we find the probability

24

of trait B being lost from the entire population to be:

$$P(loss\ of\ B) = z(1\text{-}z/k\text{+}c\text{-}cz/k)^{k}$$

We did not identify an analytical solution for the copy-if-better and copy-the-best models, and hence we solved the probability of innovation loss via simulation[3]. Figure 2 highlights how the three models are affected by the size of the sample pool, k, and the uncertainty in the payoff signal, σ, assuming that $g_A=1$, $g_B=3$, and $z=1$. In general terms, the lowest values of P(*loss of B*) are achieved by copy-if-better social learning, while copy-the-best social learning can be better or worse than random copying depending on the degree of payoff uncertainty. Payoff uncertainty is detrimental to the retention of the optimal variants under both copy-the-best and copy-if-better social learning, although the magnitude of its negative effect is much larger for the former. The relationship between k and P(*loss of B*) also varies between the three models. Random copying and copy-if-better social learning show a stable relationship (albeit the former has a lower probability of loss when $k <$ ca. 15), while in the case of copy-the-best learning we observe a continuous increase in P(*loss of B*) as function of k. In other words, random copying and copy-if-better social learning are almost unaffected by the size of the sample pool (with the latter minimally affected by payoff uncertainty), whereas copy-the-best social learning is strongly dependent on both payoff uncertainty and size of sample pool. When we allow for convergent evolution (fig.2-b), the deleterious effect of larger k on copy-the-best social learning is mitigated by the larger number of potential innovators.

---

3   Source codes are available both on the online supplement material and on the following repository: http://dx.doi.org/10.5281/zenodo.17243.

This leads to a non-linear relationship between *k* and P(*loss of B*), with the highest values of the

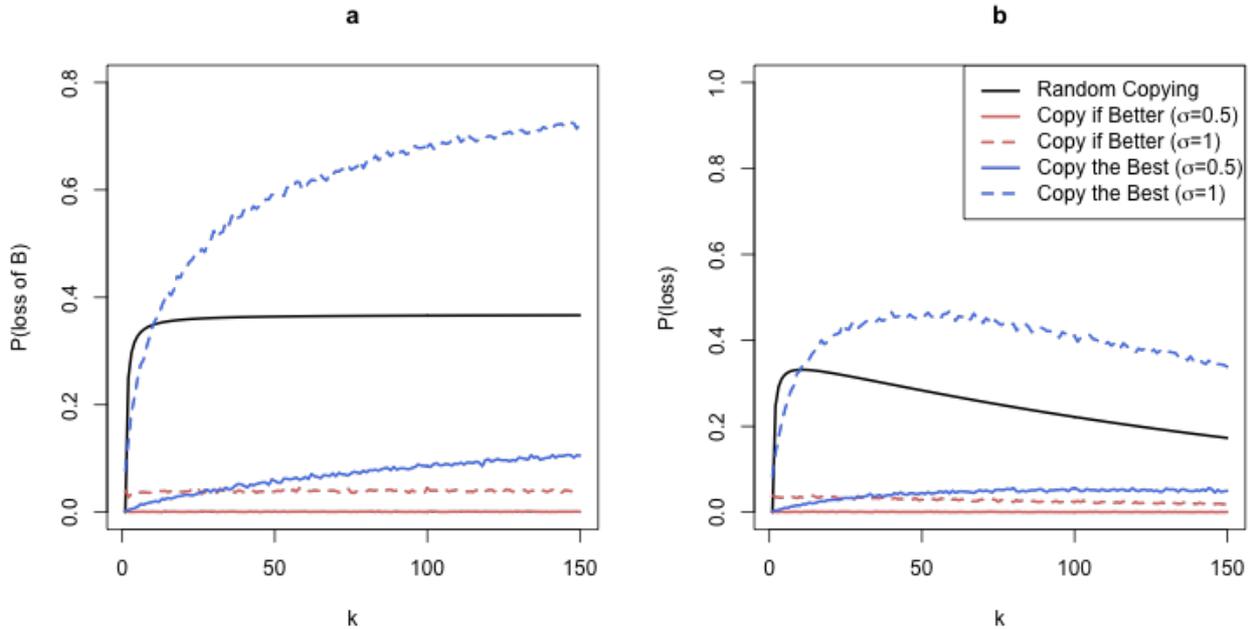latter obtained for intermediate values of the former.

 [Figure 2]

Figure 2: Probability of losing a single optimal variant (B) as a function of *k* for different models of

social learning: a) without convergent innovation; b) with convergent innovation (*c*=0.005). In all

cases g_B=3 and g_A=1. Probabilities obtained from 10,000 simulations for each value of *k*, except for

the random copying model.

Experiments 2 and 3: Spread of Innovation through a Population.   The results of the first

experiment suggest that when social learning uses a copy-the-best rule, a decreased correlation

between *g* and its corresponding payoff signal *p (i.e.* larger *σ*) can lead to smaller sample pools, *k,*

better supporting the retention of rare, beneficial traits. However, small *k* also implies a smaller

chance that a beneficial innovation occurs within the social clique (sample) in the first place. There are thus two contrasting forces: low k promotes the retention of beneficial innovations, but decreases the probability of such innovations being present in the social clique; high k increases the probability of erroneously selecting suboptimal traits present at higher frequencies, but also increases the probability that a beneficial innovation is present (and is thus available to be copied) within the social clique.

We now explore the balance between these two contrasting forces in a spatially explicit population of N randomly distributed agents, where the sample pool size is determined by the local interaction radius, defined as the k nearest neighbours to each focal agent. High values of k allow wider interaction, increasing the probability that a sampled model possesses a beneficial mutation, while low values of k produce more localised interaction with a concomitantly lower probability of innovation loss. For experiment two, we use Latin hypercube sampling (McKay et al. 1979) to ensure uniform coverage of a parameter space comprising 30,000 unique combinations for each of the three models, with σ bounded between 0 and 3, k between 1 and 500, and fixing N to 1,000, b and z to 1.0, and μ to 0.005. In all cases we initialised the entire population with $g=0$, executing 1,000 time-steps and recording the average final cultural trait value $\bar{g}$. Notice that the theoretical maximum of $\bar{g}$ is 1,000, that is the number of time-steps plus the value of g at initialisation (0).

Figure 3 shows the relationship between the final average cultural trait and the size of the sample pool (interaction radius), k, for different ranges of payoff uncertainty, σ. For all parameter settings we find that random copying is the slowest social learning strategy, with $\bar{g} \sim 5$, which is what we expect given that the rate of innovation was 0.005 and the number of time-steps 1000. When σ is close to zero, meaning that payoff uncertainty is low (or put another way, that the

correlation between the payoff signal and the underlying trait is high), copy-the-best is by far the most effective social learning rule. Under these circumstances both copy-if-better and copy-the-best social learning are positively affected by k, albeit in different ways. Copy-if-better social learning exhibits a positive correlation with sample pool size only up to $k=20$; above this threshold we do not see any increase in the rate of evolution. On the other hand, copy-the-best social learning always benefits from increasing sample pool size and the curve reaches the theoretical maximum of $\bar{g}$ when k is close to 500. Thus, when payoff uncertainty is low, larger sample pools promote the correct detection of the most successful variants and at the same time ensure the rapid diffusion of these across the entire population.

As payoff uncertainty increases, the performance of both copy-if-better and copy-the-best social learning deteriorates. In both cases there is an increased chance of accidentally adopting a suboptimal variant. However, in copy-if-better learning the sampling bias introduced by payoff uncertainty affects only the evaluation stage, and hence its effect is smaller than the in copy-the-best learning where instead the bias is also present in the selection of the social model. Indeed the difference between the rates of evolution of the two learning strategies decreases with increasing payoff uncertainty, such that once $\sigma > 1.5$, copy-if-better social learning actually yields higher $\bar{g}$ than copy-the-best learning when k is large. In other words, increasing payoff uncertainty decreases the strength of cultural selection more severely in the case of copy-the-best than it does in the case of copy-if-better. With copy-the-best social learning we observe how the beneficial effect of a larger sample pool size, k, is counterbalanced by the increased chance of inadvertently copying the trait of an individual who by chance produced a high payoff value despite possessing a suboptimal variant. As a result, when payoff uncertainty is increased, a smaller sample pool can, counter-

intuitively, promote a faster rate of cumulative cultural evolution because the benefit of minimising

the risk of accidentally adopting suboptimal cultural traits outweighs the cost of reducing the

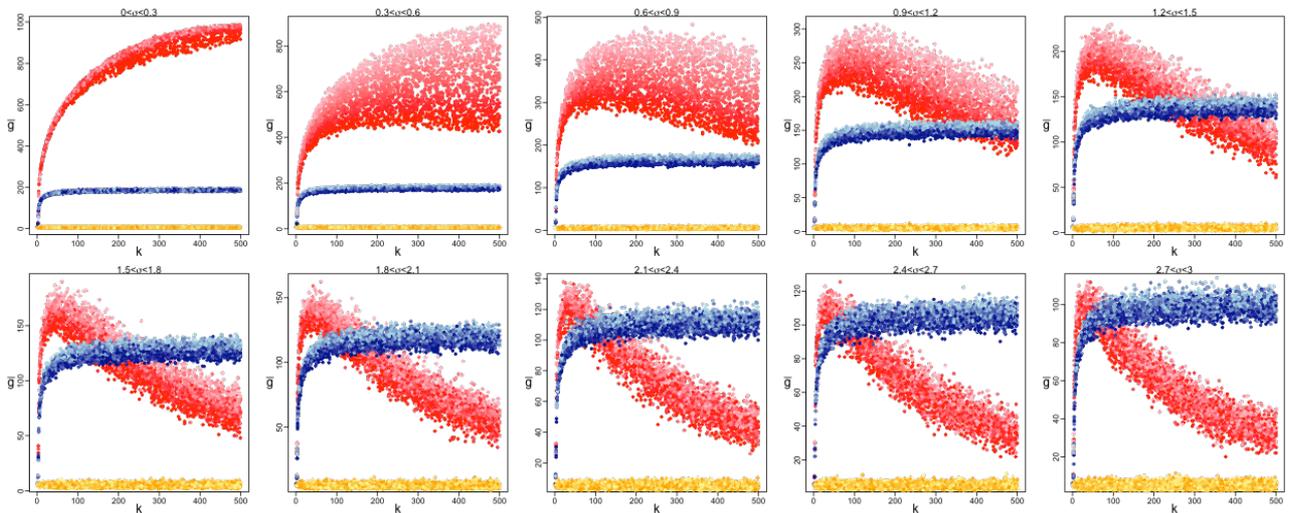potential pool of new beneficial variants.

[Figure 3]

Figure 3: Effect of payoff uncertainty (σ), and sample pool size (k) on the rate of cumulative

cultural evolution (average cultural trait value after 1,000 time-steps) for three social learning rules

(red=copy-the-best; blue=copy-if-better; yellow=random copying; darker colour indicates higher

payoff-uncertainty within each plot).

In experiment three we further explored the relationship between census population (N) and

sample pool size (k), by running the copy-the-best and copy-if-better models with four settings of σ

(0, 0.4, 1.5, and 4) and three settings of N (1000, 2000 and 4000). The results (fig.4) show that,

overall, payoff-uncertainty, σ, is still the major driver of change in the rate of cumulative cultural

evolution, $\bar{g}$, for both learning strategies. Census population size, N, is the second best predictor for

the copy-if-better model, but in the case of copy-the-best social learning the relative effect of N and

k on $\bar{g}$ is a function of σ. When payoff uncertainty is low increasing either the population size, N, or

the sample pool size , *k*, or both, increases the rate of cultural evolution. In contrast, when payoff uncertainty is high, increasing the population size always increases the rate of evolution *for any given sample pool size*, but the effect of increasing the latter can be sufficiently deleterious that it nullifies the advantage of increased population size.
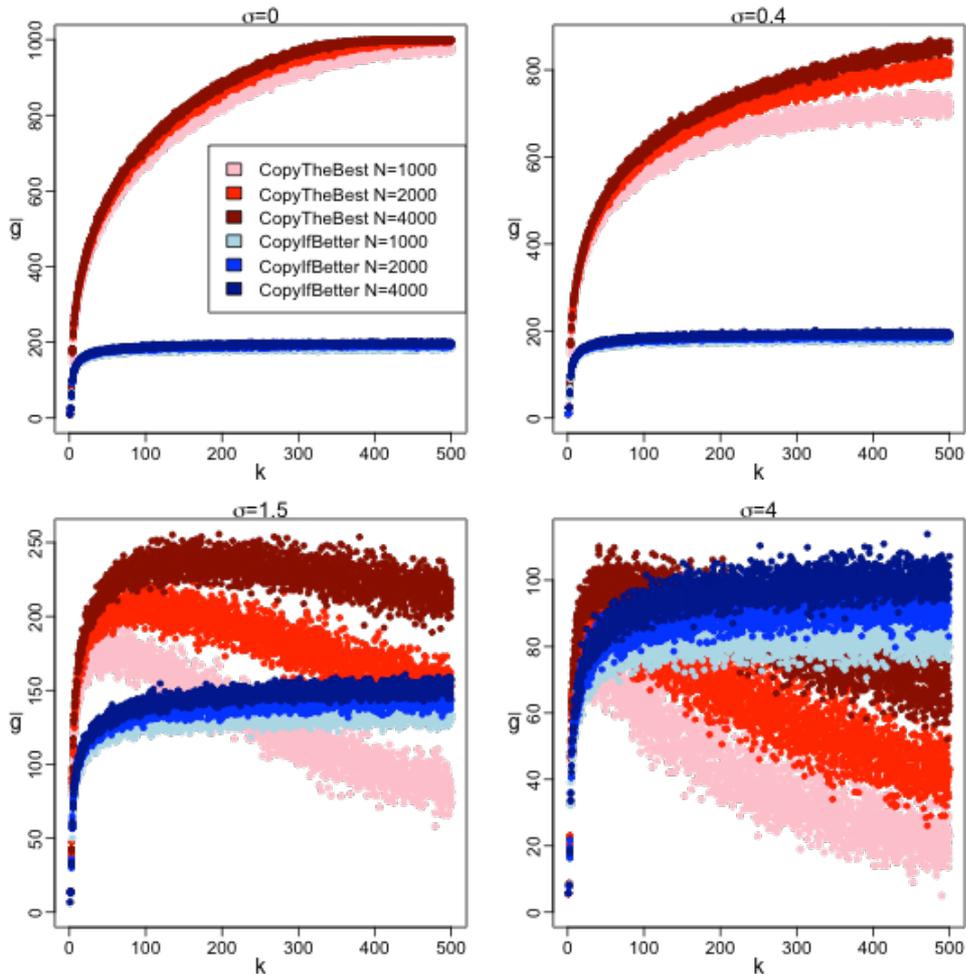
[Figure 4]

Figure 4: Effect of payoff uncertainty (σ), sample pool size (*k*) , and census population size (*N*) on the rate of cumulative cultural evolution (average cultural trait value after 1,000 time-steps) for copy-the-best and copy-if-better social learning rules.

Discussion and Conclusion

In section 1 we reviewed some of the extensive literature concerning the impact of different social learning rules and population size on cumulative cultural evolution. We then set out and reported the results of a model designed to investigate the less well explored question of how payoff uncertainty and population size/sampling might interact to promote or retard the social learning of initially rare innovations. Here we comment on the implications of our results in relation to the existing literature.

Beneficial traits often fail to spread through a population as a result of key events occurring at the earliest stage of adoption. The decision making of early adopters can potentially drive long-term evolutionary trajectories, often leading to the permanence of suboptimal traits and the loss of more beneficial innovations. This is particularly likely to occur when it is difficult to measure the value of each variant and, consequently, the comparative advantage of a novel trait is hard to evaluate (Rogers 2003). The increased learning costs of cumulative knowledge (Mesoudi 2011), a decreased investment in individual learning (Whitehead and Richerson 2009), and early chance events (Greve and Seidel 2014) can all lead to the retention of suboptimal traits and slow-down in the rate of cultural evolution. Our results suggest that a copy-the-best social learning strategy centred on the imitation of a smaller subset of highly visible, successful individuals can also

generate a decrease in the rate of cumulative cultural evolution when there is payoff uncertainty. The dynamics we demonstrate closely resemble what economists refer to as *survivorship bias* (Elton et al. 1996), a particular form of sampling bias where only highly positive payoff signals are considered, hindering the correct evaluation of the full spectrum of a cultural trait. Indeed, our simulation experiments show that when payoff is highly uncertain and the sample pool of potential cultural models is larger there is bias favouring the retention of suboptimal traits present at higher frequency. As a direct consequence, newly introduced variants, which will always initially have a lower frequency, are at an intrinsic disadvantage relative to traits that are common and whose value is measured via payoff signals that have high variance.

We noted in section 1 that payoff bias, success bias and prestige biased social learning all entail paying attention to 'payoff signals', but they differ in the likely correlation between the signal and the target trait owing to an increasing number of factors that may contribute to the generation of the payoff signal. The shape of the payoff distribution we examined here assumes an additive interaction of its generative constituents (i.e. the target trait and everything else contributing to the payoff signal), but a multiplicative interaction yielding a skewed distribution (Limpert et al. 2001) is also possible. In this scenario the correlation between the target traits and its payoff signal would be reduced and the association between the two even harder to disentangle. Although there are good reasons to believe that in many behavioural domains the payoff signal is normally distributed (Frank 2009), some empirical data on prestige and success do indeed show a positively skewed distribution, for example income (Drăgulescu and Yakovenko 2001) and number of academic citations (Gupta et al. 2005). We suspect that such asymmetric and right-tailed payoff distributions would further strengthen the effect of survivorship bias, promoting the retention of suboptimal traits

even more frequently than already observed with our model based on normally distributed payoffs. For this reason, we expect success biased and prestige biased social learning effected by copying-the-best to be even more susceptible to deleterious sampling effects than predicted by our model.

Our most striking result is that smaller, but not too small, social cliques can—perhaps counter-intuitively—promote the survival and spread of beneficial variants with copy-the-best learning strategies when payoff uncertainty is relatively high. As discussed in section 1, previous studies have demonstrated that cumulative cultural evolution requires a large enough population (Shennan 2001; Henrich 2004) or a high enough migration rate between subpopulations (Powell et al. 2009) to offset losses due to drift and the low fidelity of social learning. However, we also noted that empirical evidence does not unambiguously support this expectation (Collard et al. 2013) and our results suggest two possible reasons why increasing population size might not always be beneficial in the case of copy-the-best learning strategy. The results of our first experiment demonstrate that when there is payoff uncertainty increasing population size increases the probability of loss of a beneficial novel trait when it is very rare. To some extent this is the case for all three forms of social learning that we modelled, but is most pronounced for the copy-the-best learning rule. Although our results also demonstrate that (unsurprisingly) the deleterious sampling effect caused by the interaction between payoff uncertainty and population size can be offset by high rates of convergent innovation, we note Henrich's (2004) finding that it is harder-to-copy and/or harder-to-innovate traits that are most dependent on larger population sizes for their maintenance and/or further evolution. The results of our second and third experiments demonstrate that when payoff is uncertain, the rate of cultural evolution resulting from copy-the-best social learning is not a monotonic function of sample pool size, but is maximised by social learning within

social cliques of intermediate size. This result is broadly consistent with recent network analysis (Montanari and Saberi 2010) which found that innovations spread faster in locally connected networks than in larger well-integrated social infrastructures. The deleterious effect of larger sample pool sizes on the early stages of innovation suggests that the way in which social learners sample a population can be as or more important than the variation in the overall census population size when payoff uncertainty is high. In fact, while it is true that increasing the census population can be expected to increase the rate of evolution if the sample pool size remains unchanged, when payoff uncertainty is high it may be possible to obtain an even higher rate of evolution by reducing the population size while simultaneously reducing the sample pool size (see fig. 4). Consequently, it appears that when payoff uncertainty is high the subdivision of the census population into overlapping "cultural incubators" can increase the rate of cultural evolution as much or more than simply increasing the population size. This cautions against predicting the rate of cumulative cultural evolution on the basis of population size alone. It also suggests (contra Powell et al. 2009) that increasing connectivity between subpopulations may not always increase the rate of cumulative cultural evolution.

Finally, we note that there are circumstances in which copy-the-best social learning is expected to produce a lower rate of cumulative cultural evolution than the less-strongly directed copy-if-better learning rule. Our results demonstrate that when payoff uncertainty is low copy-the-best social learning leads to a higher rate of cultural evolution than either of the other learning rules, even as its efficacy declines with increasing sample pool size. However, as payoff uncertainty increases there comes a point where copy-if-better social learning out-performs copy-if-better learning for all except the very smallest sample pool sizes, because it is not subject to the

deleterious sampling effect that undermines copy-the-best learning as sample pool sizes increase. Empirical values for the sample pool size are difficult to estimate, although we might expect that for most part of human history it was comparatively small and close to our cognitive limits on the maintenance of stable social relationships (ca 150, Dunbar 1993). Consequently, for much of human history, *copy-the-best* social learning might have been more effective than other social learning strategies even despite its susceptibility to false inference in the face of payoff uncertainty. However, the situation might be different today, since the exponential development of information technology over the last 50 years has greatly increased the pool of potential social models (k), quite possibly creating scenarios beyond those we modelled here. Any concomitant slow-down in cultural evolution might not be visible in absolute terms (since the rate of cultural evolution could still be increasing given the larger N), but we would expect the effect of survivorship bias (and the consequent mis-evaluation of cultural traits) to become stronger, and the advantage of a copy-the-best social learning strategy to greatly decrease relative to alternative social learning strategies in the presence of high payoff uncertainty.

Literature Cited

Andersson, C., and D. Read. 2014. Group size and cultural complexity. *Nature* 511:E1.

Apesteguia, J., S. Huck, and J. Oechssler. 2007. Imitation: Theory and experimental evidence. *J. Econ. Theor.* 136:217–235.

Atkisson, C., M. J. O'Brien, and A. Mesoudi. 2012. Adult learners in a novel environment use prestige-biased social learning. *Evol. Psychol.* 10:519–537.

Baldini, R. 2012. Success-biased social learning: Cultural and evolutionary dynamics. *Theor. Popul. Biol.* 82:222–228.

Baldini, R. 2013. Two success-biased social learning strategies. *Theor. Popul. Biol.* 86:43–49.

Bandura, A. 1977. *Social Learning Theory.* Englewood Cliffs, NJ: Prentice Hall.

Bentley, A., and P. Ormerod. 2012. Agents, intelligence, and social atoms. In *Creating Consilience: Reconciling Science and the Humanities*, M. Collard and E. Slingerland, eds. Oxford: Oxford University Press, 205–222.

Bentley, R. A., M. W. Hahn, and S. J. Shennan. 2004. Random drift and culture change. *Proc. R. Soc. Lond. B Biol. Sci.* 271:1443–1450.

Bentley, R. A., and S. J. Shennan. 2005. Random copying and cultural evolution. *Science* 309:877–879.

Boyd, R., and P. J. Richerson. 1985. *Culture and the Evolutionary Process*. Chicago: University of Chicago Press.

Boyd, R., and P. J. Richerson. 1988. An evolutionary model of social learning: The effects of

    spatial and temporal variation. In *Social Learning: Psychological and Biological*

    Perspectives, T. R. Zentall and B. G. J. Galef, eds. Hillsdale, NJ: Erlbaum, 29–48.

Boyd, R., and P. J. Richerson. 1995. Why does culture increase human adaptability? *Ethol.*

    Sociobiol. 16:125–143.

Boyd, R., and P. J. Richerson. 2000. Memes: Universal acid or better mousetrap? In *Darwinizing*

    Culture: The Status of Memetics as a Science, R. Aunger, ed. Oxford: Oxford University

    Press, 143–162.

Caldwell, C. A., and R. M. Eve. 2014. Persistence of contrasting traditions in cultural evolution:

    Unpredictable payoffs generate slower rates of cultural change. *PLoS ONE* 9:e99708.

Campbell, D. T. 1974. Evolutionary epistemology. In *The Philosophy of Karl Popper*, P. A. Schlipp,

    ed. La Salle, IL: The Open Court Publishing Company, 412–463.

Collard, M., B. Buchanan, and M. J. O'Brien. 2013. Population size as an explanation for patterns

    in the paleolithic archaeological record. *Curr. Anthropol.* 54:S388–S396.

Csibra, G., and G. Gergely. 2006. Social learning and social cognition: The case for pedagogy. In

    Processes of Change in Brain and Cognitive Development: Attention and Performance XXI,

    Y. Munakata and M. H. Johnson, eds. Oxford: Oxford University Press, 249–274.

Derex, M., M.-P. Beugin, B. Godelle, et al. 2013. Experimental evidence for the influence of group

    size on cultural complexity. *Nature* 503:389–391.

Drăgulescu, A. A., and V. M. Yakovenko. 2001. Exponential and power-law probability

    distributions of wealth and income in the United Kingdom and the United States. *Physica A*

    299:213–221.

Dunbar, R. 1993. Co-evolution of neocortex size, group size and language in humans. *Behav. Brain. Sci.* 16:681–735.

Elton, E. J., M. J. Gruber, and C. R. Blake. 1996. Survivorship bias and mutual fund performance. *Rev. Financ. Stud.* 9:1097–1120.

Eriksson, K., M. Enquist, and S. Ghirlanda. 2007. Critical points in current theory of conformist social learning. *J. Evol. Psychol.* 5:67–87.

Frank, S. A. 2009. The common patterns of nature. *J. Evol. Biol.* 22:1563–1585.

Gramacy, R. B., and M. Taddy. 2010. Categorical inputs, sensitivity analysis, optimization and importance tempering with tgp version 2, an R package for treed Gaussian process models. *J. Stat. Softw.* 33:1–48. http://www.jstatsoft.org/v33/i06/.

Greve, H. R., and M.-D. L. Seidel. 2014. The thin red line between success and failure: Path dependence in the diffusion of innovative production technologies. *Strateg. Manage J.* 36:475–496. doi:10.1002/smj.2232.

Gupta, H., J. R. Campanha, and R. A. G. Pesce. 2005. Power-law distributions for the citation index of scientific publications and scientists. *Braz. J. Phys.* 35:981-986.

Henrich, J. 2001. Cultural transmission and the diffusion of innovations: Adoption dynamics indicate that biased cultural transmission is the predominate force in behavioral change. *Am. Anthropol.* 103:992–1013.

Henrich, J. 2004. Demography and cultural evolution: Why adaptive cultural processes produced maladaptive losses in Tasmania. *Am. Antiquity* 69:197–214.

Henrich, J., R. Boyd, S. Bowles et al. 2005. "Economic man" in cross-cultural perspective: Behavioral experiments in 15 small-scale societies. *Behav. Brain. Sci.* 28:795–855.

Henrich, J., and F. J. Gil-White. 2001. The evolution of prestige: Freely conferred deference as a

    mechanism for enhancing the benefits of cultural transmission. *Evol. Hum. Behav.* 22:165–

    196.

Henrich, J., and R. McElreath. 2003. The evolution of cultural evolution. *Evol. Anthropol.* 12:123–

    135.

Herrmann, E., J. Call, M. V. Hernandez-Lloreda et al. 2007. Humans have evolved specialized

    skills of social cognition: The cultural intelligence hypothesis. *Science* 317: 360–1366.

Hewlett, B., A. D. Silvestri, and C. R. Guglielmino. 2002. Semes and genes in Africa. *Curr.*

    Anthropol. 43:313–321.

Heyes, C. M., and H. C. Plotkin. 1989. Replicators and interactors in cultural evolution. In *What the*

    Philiosophy of Biology Is, M. Ruse, ed. Dordrecht: Kluwer Academic Publishers, 139–162.

Kandler, A., and K. N. Laland. 2013. Tradeoffs between the strength of conformity and number of

    conformists in variable environments. *J. Theor. Biol.* 332:191–202.

Kandler, A., and J. Steele. 2009. Innovation diffusion in time and space: Effects of social

    information and of income inequality. *Diffusion Fundamentals* 11:1–17.

Kendal, J., L.-A. Giraldeau, and K. Laland. 2009. The evolution of social learning rules: Payoff-

    biased and frequency-dependent biased transmission. *J. Theor. Biol.* 260:210–219.

Kendal, R. L., I. Coolen, Y. van Bergen et al. 2005. Trade-offs in the adaptive use of social and

    asocial learning. *Adv. Stud. Behav.* 35:333–379.

Lake, M. W., and E. R. Crema. 2012. The cultural evolution of adaptive-trait diversity when

    resources are uncertain and finite. *Adv. Complex Sys.* 15:1150013-1–1150013-19.

Laland, K. N. 2004. Social learning strategies. *Learn. Behav.* 32: 4–14.

Limpert, E., W. A. Stahel, and M. Abbt. 2001. Log-normal distributions across the sciences: Keys and clues. *BioScience* 51:341–352.

McKay, M. D., R. J. Beckman, and W. J. Conover. 1979. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* 21:239–245.

Mesoudi, A. 2011. *Cultural Evolution*. Chicago and London.: University of Chicago Press.

Mesoudi, A., L. Chang, K. Murray et al. 2014. Higher frequency of social learning in China than in the West shows cultural variation in the dynamics of cultural evolution. *Proc. R. Soc. Lond. B Biol. Sci.* 282:20142209–20142209.

Mesoudi, A., and M. J. O'Brien. 2008. The cultural transmission of Great Basin projectile point technology I: An experimental simulation. *Am. Antiquity* 73:3–28.

Mesoudi, A., and M. J. O'Brien. 2008. The cultural transmission of Great Basin projectile point technology II: An agent-based computer simulation. *Am. Antiquity* 73:627–644.

Montanari, A., and A. Saberi. 2010. The spread of innovations in social networks. *Proc. Natl. Acad. Sci. USA* 107:20196–20201.

Odling-Smee, F. J. 1983. Multiple levels in evolution: An approach to the nature-nurture issue via applied epistemology. In *Animal Models of Human Behaviour*, G. C. Davey, ed. Chichester: Wiley, 135–158.

Perreault, C. 2012. The pace of cultural evolution. *PLoS One* 7:e45150.

Powell, A., S. Shennan, and M. G. Thomas. 2009. Late Pleistocene demography and the appearance of modern human behavior. *Science* 324:1298–1301.

Powell, A., S. J. Shennan, and M. G. Thomas. 2010. Demography and variation in the accumulation

of culturally inherited skills. In *Innovation in Cultural Systems: Contributions from Evolutionary Anthropology*, M. J. O'Brien and S. J. Shennan, eds. Cambridge: MIT Press, 137–160.

Premo, L. S., and J. B. Scholnick. 2011. The spatial scale of social learning affects cultural diversity. *Am. Antiquity* 76:163–176.

R Core Team. 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Rendell, L., R. Boyd, D. Cownden et al. 2010. Why copy others? Insights from the social learning strategies tournament. *Science* 328:208–213.

Revolution Analytics and Steve Weston. 2014a. doParallel: Foreach parallel adaptor for the parallel package. R package version 1.0.8. http://CRAN.R-project.org/package=doParallel.

Revolution Analytics and S. Weston. 2014b. foreach: Foreach looping construct for R. R package version 1.4.2. http://CRAN.R-project.org/package=foreach.

Rice, T. W., and J. L. Feldman. 1997. Civic culture and democracy from Europe to America. *J. Polit.* 59:1143–1172.

Richerson, P. J., and R. Boyd. 2005. *Not By Genes Alone: How Culture Transformed Human Evolution*. Chicago: University of Chicago Press.

Richerson, P. 2013. Human evolution: Group size determines cultural complexity. *Nature* 503:351–352.

Rogers, A. 1988. Does biology constrain culture? *Am. Anthropol.* 90:819-831.

Rogers, E. 2003. *Diffusion of Innovations*. 5th ed. London: Free Press.

Schlag, K. H. 1998. Why imitate, and if so, how? A boundedly rational approach to multi-armed

bandits. *J. Econ. Theor.* 78:130–156.

Schlag, K. H. 1999. Which one should I imitate? *J. Math. Econ.* 31:493–522.

Shennan, S. 2001. Demography and cultural innovations: A model and its implications for the
emergence of modern human culture. *Camb. Archaeol. J.* 11:5–16.

Steele, J. 2009. Innovation diffusion and traveling waves. In *Pattern and Process in Cultural
Evolution*, S. Shennan, ed. London: University of California Press, 163–174.

Vaesen, K. 2012. Cumulative cultural evolution and demography. *PLoS ONE* 7:e40989.

Wakano, J. Y., and K. Aoki. 2007. Do social learning and conformist bias coevolve? Henrich and
Boyd revisited. *Theor. Popul. Biol.* 72:504–512.

Whitehead, H., and P. J. Richerson. 2009. The evolution of conformist social learning can cause
population collapse in realistically variable environments. *Evol. Hum. Behav.* 30:261–273.