Dear Editor,

We are pleased to submit a revised version of our manuscript C*ultural Incubators and Spread of Innovation* authored by myself and Mark Lake for consideration for publication in the Human Biology special issue “Approaches to Modeling Demic and Cultural Diffusion”.

We addressed all the key points and issues raised by the two reviewers, which we believe significantly improved the quality of the paper. In particular, we:

* clarified the descriptions of our models, revising our key equations;
* eliminated the “payoff-weighted” model and introduced a “copy-if-better” model which represents a more suitable intermediate stage between unbiased learning and “copy-the-best” learning.
* revised and checked the figures and their captions;
* addressed the issue pertaining to the mis-use of the term “effective population size”.

Please find below our point-by-point responses to key issues raised by the two reviewers. Original comments are typeset justified against the left margin and our responses are indented.

Yours faithfully,

Enrico Crema & Mark Lake

**Reviewer #1**

“ (Page 19) k is only mentioned in “c. Copy-the-best” but when reading on it transpired that k was affecting all of the learning biases. Include k in each description, assuming this is the case. Also, specify here whether the focal agent is included in k (it’s in a footnote later on but really should be here)”

We added a new paragraph on page 25-26 which clarifies our definition of *k,* its relationship to the existing literature, and its role in our three models. Also following the suggestion of the reviewers, *k* no longer includes the focal agent.

“page 22: I didn’t follow the results of experiment 1 at all....”

*also*

“Similarly for the probability term for Copy-with-weighted-payoff given on p.23: why does sigma not affect this probability?”

We revised our equation for the unbiased transmission and we removed our second model (“payoff-weighted model”) and instead explored a “copy-if-better” model (which does not, however, have an analytical solution). We made this change in response to one of the points raised by the second reviewer concerning the relevance of the evaluation stage in social learning (see below).

“ I don’t think the figure 4 caption applies to the actual figure 4 (the colors don’t correspond to the learning rules, contrary to the caption). I think that what is referred to in the text on p.26 as figure 5 should actually be figure 4. However I found this figure difficult to interpret, because sigma is varying both via the color and also along the x-axis. Is there a better way of illustrating this?

Figure 4 has now been removed as we no longersweeptheparameter *z* in the experimental design.. This is because we decidedconcluded that the exploration of *z* (the frequency of social learning) was a distraction from the key message of the paper, and that it was more important to to exploredifferent settings of N, following reviewer #2's suggestion.

**Reviewer #2**

“In my opinion, this paper generally confuses the size of the subset of skilled individuals that could potentially serve as a teacher for a naïve individual (the size of that subset is represented by k) and the concept of the effective size of the entire population (which should be represented by Ne).”

Our use of Ne followed that in the literature we cite by way of motivation, but reviewer #2 is correct that this is not the formal meaning to which geneticists and others are accustomed. Consequently. we rectified this throughout the paper by standardising our terminology tofocus on census population (N) and sample pool of social teachers (k). We also introduced a new experiment (experiment #3) designed to evaluate the effect of change in *N.*

“Unless I have misunderstood something, the equation presented at the bottom of page 22 is incorrect. While it works when z=1 (and it is noted that z=1 in the first experiment), it does not work for 0<z<1. This means that the equation on page 23 is probably also incorrect. The authors should double-check their math against my admittedly off the cuff calculations below.”

As mentioned earlier, we removed the payoff-weighted model, and introduced a new model (“copy-if-better”). We have now only one equation for to the unbiased model, while we obtained the probability of loss for the other models exclusively via simulation.

“A bigger problem may be that all three equations assume that the innovator is the only one in the population with the currently best cultural variant.”

We modified all equations to incorporate the effect of convergent evolution. Also, following the reviewer's suggestion the probabilities are no longer the loss of the trait for the innovator, but the loss of the trait in the entire population.

“ '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.' How does that work? Maybe it would be useful to the reader to provide 1 or 2 more sentences here to better summarize this work by Baldini. “

We extended our account of Baldini's work.

“This evaluative stage is something quite different from the other mechanisms of cultural transmission tested here. So copy-the-best, is not just different in who is targeted to learn from but also in whether or not the targeted g is ultimately adopted.”

*also*

“Page 20: It seems to me that PWF also has a selective component. Perhaps the selective component it is not as strong as CB, but it is certainly different than UB in that way, no? CB is qualitatively different from both in that it also has the evaluative stage, and this might make it difficult to compare CB to the others directly.”

Indeed we think this is an interesting point and decided that it deserves proper exploration. The new “copy-if-better” model integrates the evaluation stage but keeps the random selection of the social teacher, and hence provides a better intermediatebetween unbiased transmission (random selection / no evaluation) and copy-the-best (non-random selection/evaluation).

“Page 20: '...traits that are less beneficial can potentially...' Here, again, is a spot where the reader might scratch her head as to how a biologically selective neutral trait can be “less beneficial” than another trait.

We clarify what we meant by 'beneficial' at page 23.

“Page 22: “...The latter is measured as the median value of *g*...” Why not use the mean? Are the g values distributed non-normally at the end of the run?”

No the output is identical, but we reckon that mean is a more conventional summary statistic, hence we changed our outputs.

“Page 25: 'When σ is close to zero (that is when the correlation between the payoff signal and the underlying trait is high), panmixia promotes the immediate diffusion of beneficial traits over larger distance.' I do not see any results in the figures that have to do with distance or space. Where are the data that support this statement concerning the spread through time and over space?

Our second and third experiment is based on spatially situated agents, each with their own sample pool of social teachers defined by the *k* nearest-neighbours. Thus larger distance here refers to larger *k* and a stronger overlap of the sample pool of teachers. We added a figure to illustrate this concept and clarified what we mean by distance and panmixia in the text.

“Page 26: 'The results shown in figures 3 and 4 clearly indicate the detrimental effect of sampling bias with large *k* and the positive effect of isolation by distance*.'* […] the statement above needs to be made a bit more nuanced or the authors risk making a statement that is too strong and not entirely supported by their own results. It is interesting that UB and PWF do not show the effect of k. [...]

We specified that our result are specific to the “Copy-the-Best” model, and that this does not apply to the two other models of social learning we examined.

“ Page 29: 'The most important implication of our results is that smaller, but not too small, social cliques can, perhaps counter-intuitively, promote the survival and spread of beneficial variants.'

This is an important implication, but perhaps this is not so counterintuitive, given that smaller cliques (lower k) increase effective population size holding N, μ, z, and cultural transmission mechanism constant. This isn’t so different from the results of other studies in that sense, after all.

Our basic insight that smaller social cliques can promote the survival and spread on innovations holds, but we have partially revised our discussion of some of the earlier work (p14-18 and 25-26) to make clearer its emphasis and have more carefully nuanced our conclusions. See also next.

“ Page 29: 'Previous studies (Shennan 2001; Henrich 2004) have demonstrated that drift and low fidelity social learning can increase the accidental disappearance of existing beneficial traits and the retention of suboptimal variants in very small populations, but here we have shown that a large population size can also be detrimental to the rate of cultural evolution.' No, I disagree with the last part of this sentence. The authors did not deal with population size directly, as far as I can tell. They dealt with k directly, and that is not population size. Here the confusion between k and N and Ne comes back to affect the final conclusions of the study adversely.

As noted above, we clarified the distinction between *k* and *N*, and explored directly the relationship between the two parameters in experiment #3. The results confirm previous studies that an increase in *N* is beneficial to the rate of cultural evolution, but we also demonstrate that *k* has a crucial role, so that in certain circumstances a lower population with low *k* outperforms a higher population with high *k*. We believe that this is an important and new insight and we ilustrate it using the new figure 4.