Dear editor,

Thank you and the reviewers for contructive comments. We hope to have addressed most (or all) of them satisfactorily, and include the details below. Primarily, this required adding additional caveats, discussion and references to the Discussion, which we hope clarifies the specific contribution of this paper in a larger context.

Aside from addressing the reviewer’s comments, we note that the main model equation (1) contained an error in the original submission, and was missing a partial derivative on the final term and, somewhat related, we clarified the conditions under which our model is “well-behaved”, which involves some minor fudging of the boundary conditions (l. 136-140). These edits only have to do with the explanation of the model - not its actual implementation.

- Elie Gurarie (on behalf of all co-authors)

### Reviewer 1

#### Please highlight the limitations and strengths.

The main limitation of the paper lies with the highly abstract model. This is purely a simulation study in 1-dimension that by necessity makes severe simplifying assumptions when modelling sociality, memory, and resource tracking. Some of the modelling choices seem hard to justify - notably the social interaction range is often very large compared to the migration distance.

Our model is rightly labeled as “abstract” by the reviewer, in the sense that we attempted (really, labored) to formulate virtually the simplest possible model would show the behaviors we were interested in exploring. We note - here - that we tried about a half dozen other formulations – some of which were simpler but simply did not function (in a way, this - too - is a result, though a difficult one to formalize). Regarding the large upper limit of spatial scale of “sociality” - essentially, the largest spatial scale corresponded to a scenario where there is, (nearly) no social cohesion whatsoever, i.e. the movement is mainly governed by resource attraction and the migration memory process. Very small spatial scales of sociality are incapable of “learning” anything about the resource distribution due to highly localized information gathering, another more or less fundamental result which we do a better job of highlighting in the revised discussion.

The strengths of the paper are in its clear exposition - it uses the simple model developed to address some important and timely questions, and focuses on what are likely to be the key factors that determine migratory behaviour in real systems.

#### Please comment on the methods, results and data interpretation. If there are any objective errors, or if the conclusions are not supported, you should detail your concerns.

I believe the conclusions are in large part supported by the results, although I have some concerns about some of the terminology which I detail below. The simulation methods and analysis seem fine to me.

Thank you … see responses below.

#### Please provide your detailed review report to the editor and authors (including any comments on the Q4 Check List):

Beyond the answers above I have the following comments on the paper.

1 - the authors appear to be using a very general definition of migration. If I’m understanding their results correctly I think a lot of people would disagree with their characterisation of ‘the spontaneous emergence of a migratory behavior’. If individuals are responding to local cues is this really migration? If there is a suppression of response to local cues that would typically be associated with migration can the authors clarify how this emerges?

We modified the rather strong (and easy to misinterpret) term “spontaneous emergence of a migratory behavior” with “the emergence of a migratory behavior from an essentially resident or naive initial condition.” What we are specifically referring to here is the scenario illustrated in figure 5. This scenario is initialized with a *nearly* migratory initial condition (i.e. the distance between the seasonal ranges was 1). This slight initial condition was necessary because in our model, *some* seasonal shift must be estimated to seed the beginnings of a consistent, long-distance migration. Otherwise, we are not quite clear on the reviewer’s suggestion that response to “local cues” does not constitute migration. In our model, those pieces of information are the mechanism by which a long-distance migration is - eventually - learned. The point of the “learning to migrate” exercise is, mainly, to show how difficult (i.e. with what a narrow range of parameter values) the system can find a self-reinforcing way to expand its exploration while still retaining enough social fidelity for the population - at large - to learn to migrate.

2 - the increase in resource tracking ability due to social interactions is a well-studied phenomena, not only in animal movement but also in swarm robotics and heuristic algorithms. There’s a large literature here that is almost entirely neglected in the current work. It would be good if the authors could better place their findings relating to social taxis within the context of previous research.

We were not familiar with this rather rich and interesting literature. Unfortunately, very little of it is directly related to migration, though swarm robitic experiments could very well be used to explore the resilience and emergence of migration strategies. Nonetheless, we now mention these directions in the exploration of sociliaty and have added some references to swarm robitics (e.g. Sahin 2005 and Brambilla et al. 2013) in the revised Discussion.

3 - I was a little confused by the implementation of the reference memory. For me, it would make far more sense to think in terms of a long-term memory e.g. that changes over a decadal timescale, and a short-term memory that recalls the previous year. Instead, here the reference memory is prespecified for year zero and fixed. I also found equation 3 unclear. Is the y-subscript on kappa an exponent? Does this mean that the reference migration route is gradually forgotten? I found it hard to reconcile this equation with the results presented in sec3.5 that consisted of 30 years of simulations with kappa~0.6-0.8.

Yes, kappa is an exponent. And the reviewer is correct that the reference memory is slowly forgotten in a way that is not as satisfactory as if we had two legitimate streams of memory - one that shifted slowly and one that shifted quickly. However, implementing such a double stream would have added some additional arbitrary complications to the model. Our final working memory is, essentially, a slowly shifting mean of recent experiences (i.e. much like the decadal scale), with the reference memory contributing an additional “conservative” anchoring. In that sense, it is the resource following which is the very short-scaled (memory-less) process that is contrasted with the migration memory process.

I noticed a few other typos: L48 missing reference

fixed

L131 I’m not sure why this is a delta t’ in the integral rather than dt’

good catch, thank you.

L142 bracket misplaced for h(x,t)

fixed

L142 year sq

fixed

L155 note/not

fixed

Figures - it would be helpful to label the figures with a,b,c if using these labels in the caption.

fixed in figures

### Reviewer 2

Thank you for the mainly positive comments and the close scrutiny of the tables and figures.

l48: Reference missing.

fixed

Table 1: - kappa given twice - ˝and and˝

fixed

Figure 2: panels are not marked by letters

Figure 3: - what does the white squares mean? - results shown across five values of sigmas instead of 6

Fixed. Also added following to caption: “The white squares represents parameter combinations where the PDE could not be solved for artifactual numerical reasons, that all correspond to a failed adaptation (high mismatch).”

Figure 7: - no shades - Indicate that the scale for foraging efficiency (y axis) is rather restricted (0.6-0.7 instead of 0.0-1.0)

Fixed.

Figure 8: ˝moderiate˝ -> ˝moderate˝

Fixed