RESPONSE TO REVIEWERS

"PARENT ORIENTED TEACHER SELECTION CAUSES LANGUAGE DIVERSITY" VERSION DATED FEB 20, 2017

IBRAHIM CIMENTEPE AND HALUK O. BINGOL

1. INTRODUCTION

We would like to thank our anonymous reviewers for pointing out corrections, suggesting ideas and references. They were very useful and help us to improve readability of the manuscript.

1.1. **Organization.** This response to reviewer is organized as follows. Individual points are listed under the corresponding reviewers. Common points that are raised by multiple reviewers are addressed in Sec. 4 "Common Points".

1.2. **Markup in the manuscript.** Changes in the manuscript are highlighted and marked with section number of points. For example, changes related to the point "How k-means adapted to language", which is in section 2.2.1, is marked with 2.2.1.

1.3. **Conventions.** In order to distinguish (i) comments of the reviewers, (ii) responses given to the comments, (iii) references to the manuscript, and (iv) changes done in the manuscript, some format, font and color combinations are used.

> " This is a sample of comments of reviewer. Quotes of the comments of reviewer are given in this format. "

This is a sample response to the points raised by the reviewer. During the response, one needs to refer to a figure in the manuscript by "`Fig. 3a`". A section title in the manuscript can be given as "`Introduction`".

> " This is a sample of changes done. If one needs to quote the changes done in the manuscript, they are given in this format. "

2. REPORT TO THE REVIEWER #1

2.1. **General.**

> " The authors consider an existing model that describes how language could evolve when transferred from one generation to another and extended it to the setting in which each offspring learns their language from a particular individual who is either geographically or language-wise close to the parent of that offspring. They find that such an extended model can promote emergence of several subcommunities that share a language, as opposed to a global community sharing a universal language.
>
> I like the idea of the work. Such an extension seems natural and the findings are interesting. The paper is well structured and once the reader gets used to omnipresent typos it reads quite smoothly. I have three major comments and several minor suggestions that I list below (page X,line Y): "

Thanks.

2.2. **Major Points.**

2.2.1. *How k-means adapted to language.*

> " p3,l74: I very much like the idea of using the k-means clustering algorithm but I'd like to see in more detail how it is adapted to the language domain (possibly in an Appendix). For example, as far as I know, in the traditional k-means algorithm, at some point one takes the "centers" of each current cluster. I don't see how this is done in the adaptation which renders the findings of the paper irreproducible. "

We agree that our k-means adaptation to language domain is not given properly. As suggested, an "`Appendix A`" (page 8) is added. First k-means description in $\mathbb{R}^D$ is given. Then adaptation of k-means to language domain is described. How to handle "centers" in the domain of language is disguised in detail.

2.2.2. *Clusters vs language communities.*

> " p4,l11: While I agree that defining average within community comprehension and taking K that maximizes it is the most natural thing to do, I'd like to see some discussion of why such clusters should correspond to language communities (as in p7,l25). For example, in principle it might happen that the people with worst mutual understanding are all lumped together into one cluster, even though, logically, they are not a single language community. (This might be checkable by looking at minimum within community comprehension.) "

Thanks for pointing this important issue. The concept of inter-community comprehension is introduced in section "`2.2.1. Finding Optimum Language Clusters`" (page 3) and "`Fig. 2b`" is given and discussed in "`4.1. Subcommunities`" (page 5).
Please see Sec. 3.4 for details.
Potential problems of k-means algorithm are discussed in the "`Appendix A`" (page 8).

2.2.3. *"K^* is independent of .. N".*

> " p6,l58: I don't quite agree with the (bold) statement "K^* is independent of .. N". In Figure 3, the red data points (N=50) seem to be mostly above the green ones (N=100), which seem to be in turn mostly above the blue and purple ones (N=150,200). To me, that seems like a dependence on N, although a weak one. I'd suggest putting the statement more mildly. "

We agree that the statement "We conclude that for Models A and B, the optimum community count K^* is independent of the size of the population N" is difficult to justify. The discussion of "`Fig. 3`" in "`4.2. Number of Subcommunities`" (page 6) is revised accordingly. Now it reads:

> " (ii) There is a weak relation between K^* and N. For each model, the value of K^* is slightly less for larger values of N. "

2.2.4. *Gamma values.*

> " On a related note, what is gamma, empirically, for N=50, 100, 150, 200? "

For γ values, please see Sec. 4.1 in "Common Points".

2.3. **Minor Points.**

2.3.1. *p3,l12:*

" p3,l12: I suggest "Language evolves ⌣both⌣ through .. ⌣and⌣ as .." "

Thanks for the correction. It is done. Now, in "`2.1.2. Learning Model`" (page 3), it reads:

" Language evolves both through agents interacting with each other within a generation, and as it is transferred from one generation to the next by means of learning. "

2.3.2. *p4,l5:*

" p4,l5: This sentence might be misleading. I believe the selected partition is not necessarily the optimal one but the one found by the algorithm (which is, I agree, probably reasonably close to the optimal one). I suggest putting the sentence more mildly. "

It is true that k-means may not provide the optimum partition but near optimum is expected. The wording in "`2.2.1. Finding Optimum Language Clusters`" (page 3) is changed accordingly. Now it reads:

" For a given K, k-means algorithm provides partition $\overline{\mathbb{P}_K}$, which is expected to be close to the partition with the maximum average within community comprehension. "

2.3.3. *p5,l4:*

" p5,l4: Parameters M, S, Q are set to some specific values. Is there anything to be said about the dependence on them? "

In principle parameters M, S, Q are independent and can be selected in any way. For example, Ref [1] uses $M = S = 5$ and $Q = 1, 4, 7, 10$. Ref [2] uses $M = 20 < S = 40$.

We prefer to report the case, where there are more symbols than meanings. i.e. $M < S$. We used the values $M = 8 < S = 15$ and $Q = 4$. We also investigate $(5, 5), (5, 8), (8, 5)$ and $(8, 8)$ values for (M, S). Different sampling sizes $Q = 1, 4, 7, 10$ are tested. We found that the quality of language is low for $Q = 1$, whereas it gets higher and stays approximately the same for $Q = 4$, 7 and 10. Therefore $Q = 4$ seems to be the reasonable selection in our simulation. As previously reported, $Q = 1$ slows down learning since it is too little sampling to learn from teacher.

Since there is not much to say, no change is made in the manuscript.

2.3.4. *p5,l5:*

" p5,l5: What are the association matrices of the first generation? "

You are right that this points is missing. "`3. Model`" (page 4) section is changed accordingly. Now it reads:

" The system starts with the first generation, whose association matrices are randomly filled. Remaining generations fill their association matrices by learning. Every agent i makes exactly one child i'. The association matrix of a child is initially empty. Each child learns her language from her *teacher*, donated by t_i. "

2.3.5. *p6,l51:*

" p6,l51: The word "implies" sounds too strong to me, I'd consider "suggests" or the elsewhere frequent "indicates". "

Thanks for the correction. "Implies" is replaced by "suggests" in "`4.2. Number of Subcommunities`" (page 6). Now it reads:

" The shape of the curve in Fig. 2c suggests a power law relation between K^* and r, that is, "

2.3.6. *Highlights:*

> " I suggest "ones talk" $->$ "someone who talks", "ones that are" $->$ "someone who is", "effects" $->$ "affects" "

Thanks for the corrections. They are done.

3. REPORT TO THE REVIEWER #2

3.1. **General.**

> " This work extends a previous mathematical model to study language learning from parents to offspring, and studies two types of strategies to select teachers to communicate. It is shown that the two types of strategies induce language diversity after iterated learning across generations of learners. It also finds a power-law relation between the optimal comprehension level and the size of imitation set for selecting teachers. The mathematical parts of the model are solid, and the results are convincing (though see below). It is a topic suitable for JTB. However, there are several concerns that need to be addressed before this simple work can be published. "

Thanks.

3.2. **Gong & Shuai 2012 Proceeding B.**

> " First, in the background part, the authors only reviewed previous mathematical models from Nowak's group and briefly mentioned some behavioral models of iterated learning and similar laboratory experiments. However, the idea of learning to people that have high comprehension score during cross-generational learning has been implemented in some behavioral models, such as Gong & Shuai 2012 Proceeding B. In that work, cultural transmission (selection) is based on individuals' linguistic understandability. This work should be mentioned, since the idea of the current paper is not new. "

Thank for pointing out this missing reference. "`1. Introduction`" (page 1) is revised accordingly. Now it reads:

> " Yet another group consider language as a means to transfer meanings between individuals through signaling structures [8-10]. "

where Ref [9] is Gong and Shuai 2012.

3.3. **Comprehension scores of both models are already above 0.125.**

> " Second, in the result part, there are some confusing points. For example, on page 5, above "Subcommunities", the authors said: "In Fig. 1, comprehensions of both Models A and B are below the threshold of $W_r(\mathcal{P}) = 0.125$ for $r < 0.05$". However, it is quite clear in that Fig. 1, except under a small $r < 0.05$, comprehension scores of both models are already above 0.125. Please specify this. "

We agree that wording was not good. It is changed accordingly in "`4. Results and Discussion`" (page 5). Now it reads:

> " In Fig. 1, comprehensions of both Models A and B are below the threshold of $W_r(\mathcal{P}) = 0.125$ for $r < 0.05$. The comprehensions increase slowly and reach to the level of Model-C as r approaches to 0.5. "

3.4. **Diversity.**

> " In addition, since the authors talk about "diversity", only showing within community comprehension score is insufficient, you need to also show comprehension score between communities, in this way, people can really have a sense that the system is diverge! Showing the average comprehension score of all communities is also not enough, because it is possible

that the low average score is simply due to low score of some community, and between communities, comprehension score is still high; in other words, every community is trying to develop the common language, but some communities are delayed, so the average score of all communities is not high. This is certainly not divergent! "

We agree that our presentation was focused on within community comprehension. We have revised the material and included discussion of inter-community comprehension, too. Inter-community comprehension $I(\mathcal{C}_\alpha \leftrightarrow \mathcal{C}_\beta)$ and average inter-community comprehension $I(\mathbb{P}_K)$ are defined in "`2.2.1. Finding Optimum Language Clusters`" (page 3). As suggested, average inter-community comprehension are given in "`Fig. 2b`" and discussed in "`4.1. Subcommunities`" (page 5).

Please also see Sec. 2.2.2.

3.5. Single teacher.

" Furthermore, the setting that each child only learns from a single parent throughout the learning process is not realistic at all. Teachers should be chosen proportionally and there must be more than one learning cycle. This actually introduces some noise, because in this process, some individual from the imitation set having low comprehension score can still be chosen, which may delay or even reverse the divergence process. Since most parts of the model are already based on probabilities, the learning part should also be like this, simply by setting more than one learning cycle for each child and let teachers in those cycles be chosen probabilistically. "

We agree that having single teacher is too much abstraction but an effective one. Effects of having many teachers can be investigated as it is stated in the "`4.3. Future Work`" (page 7) section. But we feel that it is not in the scope of this paper. Now it reads:

" For example, evolution of language can be perceived as a cultural process where some group of people is responsible for the transfer. That is, more than one teacher could be assigned to each child. "

3.6. Power law.

" Finally, in order to show that the relation between K and r is power-law, why not draw Fig. 3 in a log-log scale to see if there are straight lines or not. In addition, some more discussion on this point is needed. "

Please see Sec. 4.1 in "Common Points".

3.7. Cross-generational cultural transmission.

" Third, in the future work, it would be interesting to further study if grandparent grand-child interactions can help keep language from divergence. This can be simply implemented by keeping the grandparent's status and let child have a chance to learning from them, following exactly the same setting for him/her to learn from parent generation. Some early work like Gong & Shuai 2015 book chapter: Simulating the effects of cross-generational cultural transmission on language change has addressed this based on a behavioral model. A mathematical model like the one in the current paper may complement the findings. "

We agree that, although it is interesting, we do not consider the impact of grandparents and their generation in this work. Probably one should consider this together with multi-teacher case of Sec. 3.5 as a future study. We update the "`4.3. Future Work`" (page 7) section. It states the possibility of teachers from the generation of grandparents. Now it reads:

" Once many teachers case is considered, one may also consider teachers not only from the parent's generation but the generation of grandparents, too [20]. "

where Ref [20] is Gong and Shuai 2015.

3.8. Some tiny points:

" Some tiny points:
    1) page 7, right column, line 6: "effect" should be "affect"
    2) page 7, right column, line 14: "which we communicate" should be "whom we communicate" "

Thanks for the corrections. They are done.

4. Common Points

Common points that are raised by both reviewers are addressed here.

4.1. Power law discussion.
Issues related to power law are raised in Sec. 2.2.4 and Sec. 3.6. Section "4.2. Number of Subcommunities" (page 6) in the manuscript is modified to enhance power law discussion. Now "Fig. 3" has the same values in both linear-linear and log-log scale. Discussion of straight-line patterns in log-log plot is included in "4.2. Number of Sub communities" (page 6). γ values of Model-A and Model-B for various N can be seen in Fig. A1. These values are given as "Table. 1" in the manuscript.
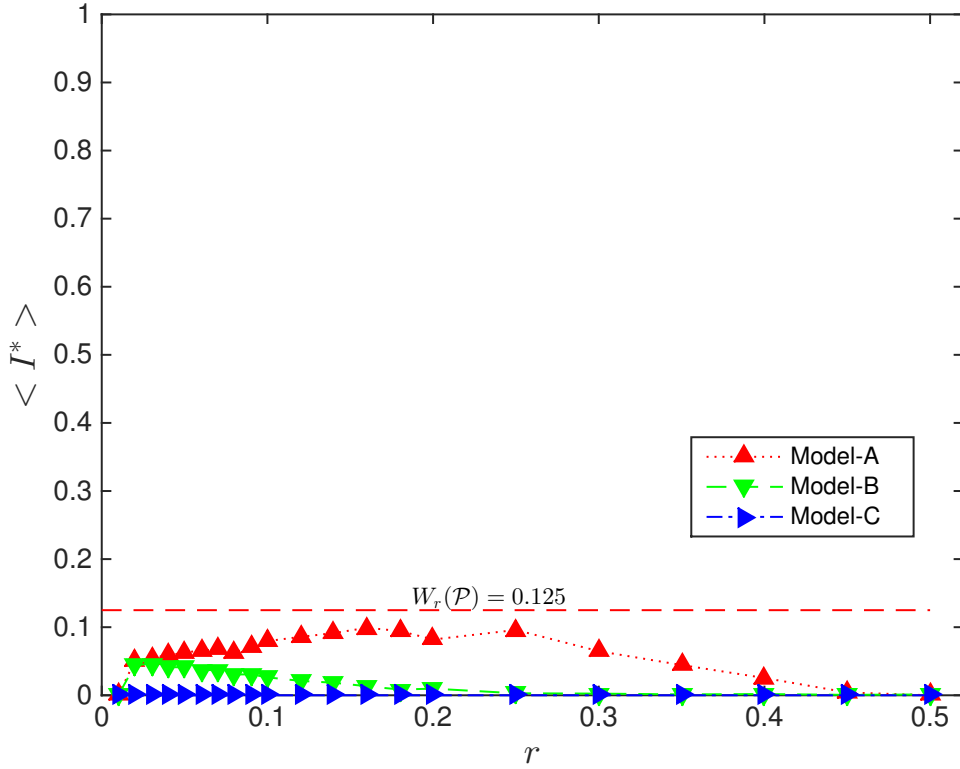

FIGURE A1. γ values of Model-A and Model-B for various N.

References

[1] M. A. Nowak, J. B. Plotkin, and D. C. Krakauer, "The evolutionary language game," *Journal of Theoretical Biology*, vol. 200, no. 2, pp. 147–162, 1999.

[2] M. A. Nowak and D. C. Krakauer, "The evolution of language," *Proceedings of the National Academy of Sciences*, vol. 96, no. 14, pp. 8028–8033, 1999.