Dear Dr. Brenda McCowan and Reviewers

We truly appreciate for your thoughtful and constructive comments on our manuscript.
We have submitted the revised version of “Behavioural synchronization in a multilevel society of feral horses”
We have uploaded the revised version in BioRxiv (https://doi.org/10.1101/2021.02.21.432190).

We wrote responses in blue texts.

Thank you very much.

Sincerely yours,
Tamao Maeda

Corresponding Author
Tamao Maeda
Wildlife Research Centre, Kyoto University
Email: maeda.tamao.76a@st.kyoto-u.ac.jp

****

Reviewer1

In this study, the authors investigate the effects of social organization and social network connectedness on behavioral synchronization of individual animals. On free-ranging horses that show multi-level societies, they compare empirical data and model simulations to reveal interesting associations between social units, social interactions, and the propensities for animals to synchronize resting and movement states both within and across social units. The study provides novel insights into the movement ecology of multi-level societies, which remain largely understudied despite a recent wave of analytical work. Aside from its novelty, I found the manuscript to be well-written, with the authors also making informed decisions regarding their design and implementation of model simulations. I nevertheless have a series of (mostly minor, but some major) concerns and comments I feel that the authors should address prior to deeming this work fit for publication:

Thank you very much!

Comments

Abstract

I feel that the summary is missing a final line of discussion, on the biological relevance of the authors' findings. Specifically, what does your finding that behavioral synchronization occurred at the intra-and inter-unit level mean for our understanding of horse social structure, or more generally of multi-level societies?

We added the discussion part in the summary (L24-25).

Introduction:

P2 L29: Provide a definition of "synchronization" here, before you emphasize why it is important (as you do in the rest of the paragraph). I come away from this paragraph knowing that it is important to study synchronization of behavior, but not necessarily what it actually means.

We added the definition of the behavioural synchronization as suggested (L29-31).

P2 L37: Here I feel that the authors could add a short paragraph, may be 2-3 lines, on what we know about the socioecological or biological importance of behavioral synchronization. For instance, is synchronization important to cohesion, social stability, resource finding or access, or predator avoidance? Is it linked to fitness and survival (other such advantages/benefits)? If there are any studies that, in addition to just examining evidence (or lack thereof) for behavioral synchronization, get at the above links, then review them here. Alternatively, review them within the follow-up paragraph in which you focus on reviewing the evidence for behavioral synchronization in single-layered groups/species.

We added some explanations on the ecological and social function of synchrony(L40-44).

P3 L71-75: These two sentences don't quite make sense, and need to be re-worded. Firstly, do you say "the behavioral synchronization", or just "behavioral synchronization"? The second sentence could be:

"Although the mechanisms underlying behavioural synchronization remain unclear, they are important to study and unravel in multi-level societies in order to better understand the collective features of such societies." Related to this sentence, are you actually trying to

understand the mechanisms (causal factors) of synchronization here, or just whether or not there is within-versus between-unit synchronization in the first place? If, as I believe, it is the latter, then perhaps don't even use the word "mechanisms" here. I would just say something like: "Whether multi-level societies show behavioural synchronization remains unclear, but it is important to address this question in order to better understand the collective features of such societies."

We changed as suggested.

P3 L79-80: Re-word for better clarity as: "..., which showed a two layered structure of units (DEFINE) nested within a herd (DEFINE)".

We changed as suggested.

P3 L82-83: Again, I am left wondering as to what exactly you mean by "a model of behavioral synchronization", or specifically "synchronize their behaviors" here. It would be good to re-state a definition for "behavioral synchronization" here.

We changed like followings: In the current study, we further apply this data collection to investigate whether horse multilevel society shows synchrony in resting/moving timing, and if so, whether the extent of synchronization changes within and across units(L89-92)

Methods:

P4 L123: Could you better define how you identified and distinguished a 'young individual'? Also provide more details regarding how you distinguished between males and females. Did you use any of the physical features that you list in line 114 for this?

We added the definition of the age class (L128-129) and males/females (L126-127).

P6 L191: Here I feel that a lay-reader would benefit from having an overall description of the model up front, rather than further down under Model Design. After you outline your overall goal, it would help to have 2-3 lines describing what type of model you used to achieve this goal, and what the model does to achieve the same. In other words, perhaps move the first 2-3 lines under the Model Design sub-section to up here, and also state following this line that you used (for example) this general model design to test specific hypotheses pertaining to the effects of social organization on the synchronization of behaviors (you don't have to state each individual hypothesis here, but give just the big

picture). I feel that this would help the reader better follow the overall purpose of constructing these models.

As suggested, we moved model design first and added a short description on our aim in the beginning.

P7 L205-207: Is this a fair, and a common, assumption to make, considering the large population sizes and the cohesivity (or lack thereof) of the social organizations that we are dealing with here? For instance, would this “awareness” not depend on the network connectedness of animals, i.e. differ in accordance with whether or not an animal was part of your own sub-group? Or on ecological visual barriers like trees, unless we are to assume that there were very few of these...? It may well be that I am missing something here, but could you better justify or explain this assumption, perhaps also cite previous work that has also made this assumption?

We deleted this part, because this description is partly wrong. Agents do not have a global view in independent hypothesis and unit-level social hypothesis. We rather checked whether they only see the behaviour of nearby individuals or they have (nearly) global view using the model. We considered this should be rather discussed in the discussion part (We added some discussions on this problems, for instance L481-482).

As for ecological barriers, in our case, the large part of the landscape had almost no tree or other obstacles. We added this description in the method (L329-330). We observed horses respond to horses or free-ranging cows appeared more than 800m apart several times (almost same as the distant from an edge of the field to the opposite), so we presume they actually have wide range of vision or hearing.

As a result, the model D selected, which means that individuals were aware of other units’ behaviour, but more attentive to same unit members (which means individuals nearby). Indeed, as written in the discussion, it is possible that units only respond to the nearby units, but we could not test this because there was technical limitation to implement the range of the view (please see the Limit of this study also).

Meanwhile, even if horses respond to a part of units, it may not largely change the results in 30 minutes interval. $\Delta T_{01,s}$ (the phase latency) is 50 and 25 minutes. For example, after $P_{m:r}$ started, more than a half of population should be already resting and another 30

minutes later all the population should have already rested. Even if units only react to the nearby units, such effect may become minor in 30 minutes window.

We hope this would answer your question. We would appreciate if you have further comments.

P7 L215-218: From this, I gather that information regarding animals' changing states and social behavior were input in the model. By "social behavior", do you mean their social network connectedness? Given that you identified individuals' sex, would you also expect sex-based differences? If so, and more generally, is it possible to incorporate individual attributes (sex, age, but also social attributes like rank, node-level social connectedness or centrality, or size of the unit or community) in addition to, or instead of, dyadic attributes in your models? More generally, would asking whether individuals of certain attributes sociodemographic characteristics and attributes have a greater-than-expected tendency to synchronize their behaviors be an interesting question to ask in the first place? If this is difficult to pull off given the scope of these models, then perhaps the authors could at least speculate (in the Discussion) on how individual and attributes and characteristics might influence their propensities (or lack thereof) to synchronize their behaviors.

We added the discussion (paragraph L373-377, L446-L467).

The network contains the difference in connectedness and centrality, so it already considered in the model. As Reviewer 3 suggested, it may be interesting if we could create another type of network (like grooming), but different unit members rarely conduct such direct social behaviours.

As for the sex, intra-unit level dominance rank, age and personality, we presume it would not affect the synchronization in herd-level because each unit contains individuals with different sex, rank, etc and intra-unit level synchronization is much stronger. For instance, even if a bold individual tends to initiate the movement in intra-unit level, an individual which may move next is a shy individual in the same unit, not bold individual in a different unit.

P7 L222, 225: Supplementary (material?)

We specified the section we refer to.

P8 L234-237: This line doesn't seem to make sense. Consider re-wording for better clarity. Did you mean to say "irrespective of the individual"?

Yes. We changed as suggested.

P10 L308: I reckon this should be "model fit", not "fitness of models".

Done

P10 L311: Change to "each 30-minute window".

Done

P11 L315: Could you better justify why you used a non-parametric (K-S) test here? I guess it was it because of the inter-dependencies of dyadic data, but isn't that why you used Mantel tests? More generally, why use both K-S tests and Mantel tests? Perhaps I am missing something here...

Mantel test uses the correlations of two matrix. So for example, two normal distribution $N(x_1, \delta)$ and $N(x_2, \delta)$ should have correlation 1.0, although the peak of the distribution (x_1 and x_2) is different. In this data, it is also important to evaluate the similarity of the peak position, so we used K-S test.

We added the explanation like:

"We used Mantel test to evaluate the similarity of synchronization rate matrix as a whole, especially the ratio of intra- and inter-unit synchronization rate." (L345)

P11 L327-329: Could you better explain what 'score' was used to determine the extent of deviance from the null? I am guessing that this 'score' is simply, as you state in the previous line, the proportion of models that showed better results, than the null, in terms of being able to predict the observed data. If this is correct, please explain this better in the last line, otherwise provide more details here on how you calculated this 'score'.

We revised the sentence and clarify the definition of the score (L354-).

Results:

P11 L346: This first line is probably not necessary, since we know by know as to what these models function for.

We deleted the sentence as suggested.

Discussion:

P12 L370: I would also be clear and say “synchronization of behaviors” here. So perhaps say “...dynamics of behavioural states, specifically the synchronization of resting versus movement,...”

Done

P13 L417: See my comment regarding P7 L215-218, and thoughts on how the authors might provide a short paragraph here speculating on how inter-individual differences in attributes (e.g. age, sex, rank, centrality) might affect behavioral synchronization.

We added the paragraph. In short, it is not known how attributes can affect intra-unit level synchrony, but there is almost no study in inter-unit level (L446-L467). Please also see our answer to P7 L205-207.

P14 L418: See my comment regarding your assumption on P7 L205-207.I understand that the simplicity of your model would make it broadly applicable. That said, could you say something here about whether the (key) assumption you made regarding animals’ awareness of the states of other animals at any given time is alsobroadly applicable? For instance, how would that differ across, say, horses that presumably live in open plains, and nonhuman primates living in multilevel societies that inhabit tropical and temperate forests where the vegetation is more dense?

We added some discussions here (L468-).

Reviewer2

This study compares an empirical data set on a feral population of horses with simulated data from multiple agent based models to ask whether behavioral synchronization in a multilevel society can be described by a simple model, and whether this model must include the network of social relationships to accurately describe the pattern of synchronization observed. I think the study is well-designed, the hypotheses behind the different models make sense, and the measures of behavioral synchrony (e.g., number of individuals switching from one state to the other) are appropriate. Some aspects of the sampling process for the empirical data could be clarified a little more (see specific line comments below). For instance, behavioral data were obtained using aerial drones which took photo or video scans

of the entire herd, and it is unclear whether a validation process was done to verify that e.g., a horse described as 'resting' in the scan image was truly resting according to an observer on the ground. I really like the inclusion of a social network into the agent-based models to investigate the potential influence of social relationships on behavioral synchrony. I think those familiar with network analysis would like to see additional information about these networks, such as density of the association networks generated for each scan observation or how many components were in each network (if not a single connected component). Finally, there are grammatical errors throughout the paper that cause some confusion about the authors' meaning, particularly in the Discussion section, which could benefit from revision. I have made some suggested revisions in the line comments below, but this was not possible for all such occurrences.

Thank you very much!

We added some information on the networks and mentioned reliability tests of behaviour identification from drones (supplementary the reliability test)

Line 120: Please explain further the definition and calculation of 'zones'. Because I am unfamiliar with such analyses, I am a little concerned that when the authors say that they calculated "…distances between all pairs of individuals in the same zone" there are some pairs that occur in different 'zones' yet their inter-individual distance would still be relevant to calculate. A definition of zones could potentially clarify this.

We added the definition of zones. We separated like this, because those two flat areas were separated by rocky hills because they are visually divided, and the migration between zones rarely occurred during daytime (L114)

Line 121: Were different times of day over-sampled relative to other times of day? Although the Supplementary files show that an analysis was done to see if horses' behavior differed according to time of day, the additional information about level of sampling per 30-minute interval would also provide helpful context.

Yes, there was a variation in sampling number each day. We added the information in Supplementary. We considered each observation to be independent because at the unit-level, 8 min were enough for them to change the association (Christensen et al. 2002), and horses moved 44.1 [30.5, 93.7] m (median [1st, 3rd quartile]) in 30 min, which should be enough to change unit-level association as well.

It is still possible to make a bias because the unit availability changes, but the result did not change when we used a network made from the part of the observations (please see the discussion with reviewer 3), so we presume that the bias do not affect the simulation significantly.

Line 131-132: I don't quite understand what was done with the histogram of the inter-individual distance data. What is meant by "…from the distance data to the shape under the R environment"? My guess is this examination of the data was performed in R or RStudio (the 'R environment'), but it seems like some words are missing from this sentence.
Yes, we wanted to say that the histogram was made in R. We changed the wording according to your advice.

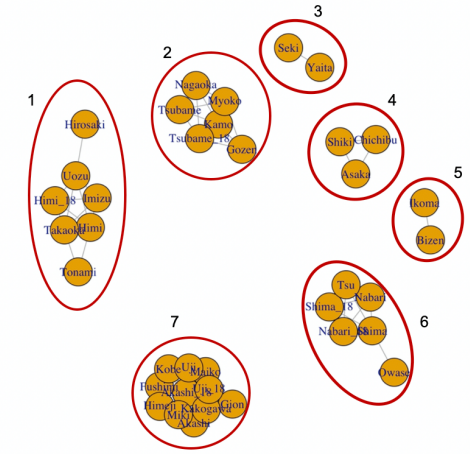
Lines 133-135: Suggest revising to read: "As shown in Figure 2, the histogram of inter-individual distances had two peaks – at the 2nd bin (0.9-1.8m) and at the 55th bin (49.7-50.6m) with a bin-width of 0.92m. The minimum frequency, or nadir, between these two peaks was observed at the 12th bin (10.1-11.0m), and we selected this as the threshold distance between…."
We changed as suggested.

Lines 140-142: The word 'scaled' doesn't seem quite right here. Perhaps 'assigned'? Pairs of horses whose inter-individual distance was smaller than 11 meters were assigned an edge weight of 1, based upon the threshold distance defined above. All other pairs were assigned an edge weight of 0.
We revised the paragraph to make it clear.

Line 143: What is an 'unconnected component'? Is this a typo and the authors meant to write 'connected component' or is there another concept being applied here? Also, the authors say that pairs of individuals were considered to be associated when they were part of the same unconnected component. What is the different between being 'associated' and identifying 'social relationships'? Is being 'associated' only needed to determine which individuals in the herd were isolates in the network graph, and it should not be conflated with what defines an edge in the network?
Following figure has done in the individual-level with foals, but the idea is the same.

It means that when they are connected directly or even indirectly, they are considered as a same temporary group (each circle in the figure) = associated in the network. This procedure has done to conduct data stream permutation in a different paper actually.

We revised the paragraph to make it clear (L160-).



Line 156: Shouldn't this say "…in our case, a scan every 30 minutes…"? Because a single scan was not 30 minutes in duration, but spaced 30 minutes apart. The duration of the (drone) scans is noted to be 4 minutes and 24 seconds on average.

Yes, we changed as suggested.

Lines 115-118 and 159-164: Did the authors validate that their assignment of resting vs. moving, as captured in the drone photographs, was correct?

Yes, we added the section in Supplementary Appendix (Reliability test of behavioural identification from drones). Just to mention, we used 2019 data for the reliability test, not 2018 data we actually used for the analysis.

Line 159: How many of the scans/ observations did not include 90% of the units (and were thus omitted from analysis)?

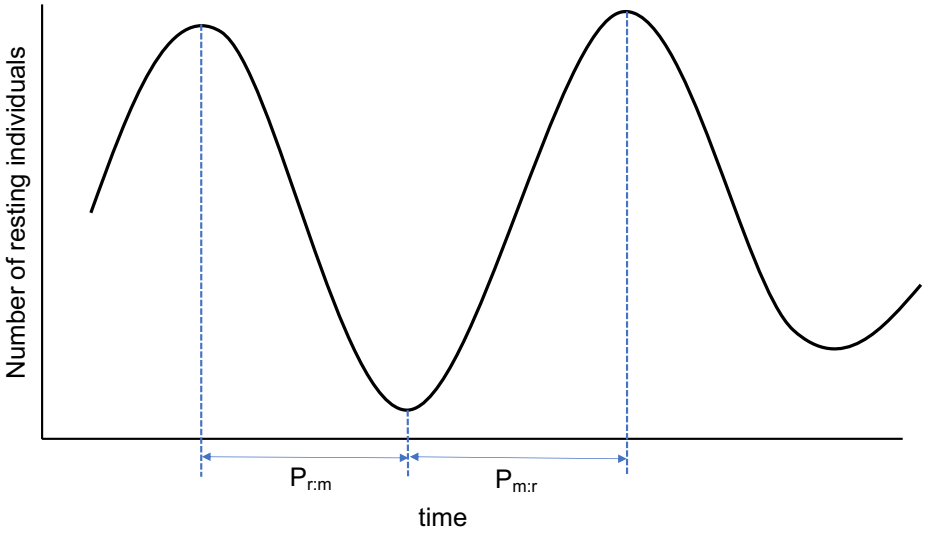
10 out of 19 days did not. Please also see Figure S3.

We forgot to mention the reason we did not include the July 5th data in the last version, so we also added it in the method section.

Lines 164-167: I don't fully understand how these phases were defined. Please revise these

sentences.

We added figure 4 to explain.



Line 178: Please define refractory period for readers.

We added the definition with citation (L199-201).

Line 310: Revise to read: "…how many individuals changed state…"

Done

Line 313: Revise to read: "…using linear regression in the R environment."

Done

Lines 321-322: Revise to read: "Mantel tests were performed using the R package 'vegan', and K-S tests were performed using the function 'ks.test' in the R environment."

Done

Lines 323-324: Revise grammar; for instance: "…horses live in a multilevel society and are therefore expected to show social cohesion and behavioral synchronization."

Done

Lines 355-359: Revise grammar, for instance: "Figure 6 shows the comparison between model-generated synchronization scores and synchronization scores from the empirical data.

The model simulations that did not consider social relationships (i.e., independent, absolute anonymous, and proportional anonymous models) showed a lot of overlap in the histograms of intra-unit and inter-unit synchronization scores, unlike the observed data which show clear separation between intra and inter-unit synchronization scores (Figure 6)." In addition, it would help to better incorporate Mantel test and K-S test results (in Table 2) at the appropriate comparisons in the text to better integrate Table 2 information with results descriptions.

We revised as suggested. Also, we corrected the name of models in Table 2

Lines 370-372: Maybe rephrase a little here. Perhaps something like "…better matched the empirical data than the null model (model A), whereas…" The phrasing 'showed better results' is a little imprecise.

Done

Line 385: I assume you are comparing the observational data to the model simulation data when you say "The observation data had higher intra- and inter-unit synchronization…" but this should be made clear, and if so, are you referring to ALL models? However, I am not sure what is being compared in the following lines (i.e., "number of individuals changes state…" is higher than what?)

We are referring to model Da and Db (as Ca and Cb had higher intra-unit synchronization rate than observed model). We added some descriptions.

Lines 385-400: Grammatical revisions are needed to clarify what discussion points and conclusions are being made. This paragraph is confusing as it is currently written.

We revised the paragraph to make it clearer (Paragraph L417-).

Line 406: Could this not have been examined in the agent based models? Can the parameter settings, or the design of the model for agents' behavior, be changed to allow the agents to 'see' only the other units that are closest to them, such as implementing a certain distance radius around an agent?

Yes, it is theoretically possible, but it was technically difficult for our data because we also need to model the movement of the horses to do that (please also see Limitation of the Study after Discussion). Especially, when foraging together, the herd should show collective behaviour, so we cannot set like horses move independently. In the current sparse time scale,

it was difficult for us to make movement model.

We added some discussions

Reviewer3

The authors propose a very interesting approach using a multi-agent model tested on several underlying social networks in order to identify the social structure that is most likely driving synchronization behaviour in feral horses. The hypotheses taken to generate social networks and the different aspects of the model are well grounded as well as the data anlysis conducted to analyse the correspondence between the models results and the empirical data. Those latter, based on drone observations are particularly promissing (despite the "low" autonomy of the drone) and the results are quite convincing.

The few remarks (below) I have are rather a consequence the inspirations i had reading the paper, rather than real criticisms about the approach and the methodology that are very convincing to me.

Thank you very much!

A first remark concerns the term synchronization, as having in mind models like the synchronization of fireflies it is maybe a bit misleading and the current case corresponds rather to a more global case of social influence. But if the term is the real one used in this particular field, I would totally accept it could be a seen as a synchronized process at a larger time scale.

According to a review by Duranton and Gaunet (2016), the term “behavioural synchronization” covers a very large field. We see synchronization is used also when the social process is involved. In addition, it is also used for a large time scaled behaviour like resting/moving and even reproductive synchrony. So, we would like to keep using synchronization in this paper.

A second one is rather a perspective about the global method concerning the generation of social networks that is quite classical in the field of animal networks study. The point is that using spatial distance to generate networks you cannot really distinguish in between a spatial influence and a social one. This is not totally the case there as the different social networks tested enable to fill a gap in between the two options. However, one possible solution (and it

is maybe the one tested there but it is not totally clear) would be to use half of the dataset in order to generate the social network and once generated to use the other half in order to test the different hypotheses. Such option would be in particular interesting when using directed relationships rather than bidirectional ones, trying to capture a single sided influence especially introducing the perception of the individuals that is oriented (therefore a symetric relation would correspond to the case where two horses are standing face to face).

A complementary approach that could potentially shed a different light on the system concerns the analysis of the evolving social network in particular concerning the evolution of the social communities identified along time.

Indeed, it should be interesting to compare it with different networks. Unfortunately, in our case, we do not have a social interaction data across units, because direct interactions, mainly greetings and fighting between males, are pretty rare actually.

We have tried separating the network as you suggested.

In original model:

We have 19 days observations -> created network (original network)

We selected 8 days of them, which had 21 or more units-> obtained other parameters

We separated the network like following,

8 days used to obtain parameters -> network 1: "day used" in the table

Remaining 11 days -> network 2: "day not used"

As a result, they have some variations in the results, but they were mostly consistent with the original models and the models with different parameters (different mimetic coefficient or refractory time period).

We may be able to say that the simulations were somewhat robust, but we are not sure if these results could be used to test the hypothesis you mentioned. We would appreciate if you could give us further comments.

	K-S test (Δn_m, $P_{r:m}$)	K-S test (Δn_r, $P_{m:r}$)	K-S test	Mantel test (sync rate)

model	parameter type	D	eval	r	eval	D (sync rate of intra-unit)	eval	r	eval	score
	day used	0.74	+	0.72	+	0.84	+	0.62	+	1.00
	day not used	0.73	+	0.73	+	0.80	+	0.61	+	1.00
Da	original model	0.54	+	0.58	+	0.92	+	0.60	+	1.00
	max*	0.77		0.77		0.99		0.60		
	min*	0.32		0.33		0.84		0.32		
	day used	0.86	+	0.85	+	0.66	+	0.59	+	1.00
	day not used	0.85	+	0.84	+	0.66	+	0.60	+	1.00
Db	original model	0.70	+	0.71	+	0.71	+	0.63	+	1.00
	max*	0.82		0.84		0.87		0.65		
	min*	0.48		0.49		0.61		0.58		

*max and min is the maximum and minimum value of D or r in models with parameter set 1-9 (supplementary appendix Table S2 and S3).