Manuscript 70741v1

Title: Prior experience of captivity affects behavioural responses to 'novel' environments

1. Basic Reporting

In general, the overall structure and flow is okay. However, some of the writing is not clear and can be frustrating to find the necessary information. I am also confused and lost with some of the results and statements, some of them were contradicted with each other. The formatting can be improved as well. Please see the details below.

Introduction: Overall is too long, please try to trim down a bit, especially L49-69.

L38, 109, 432, 451: Please discuss the arguments, do not just use "but" or "but see" for those references. This is inappropriate. Not every reader is an expert in the field and remember every single piece of work.

L51: "many studies use a two-step approach", but only one reference provided. Add one or two to support the statement.

L121-123: It is unclear here, why and what social responses have been used for this particular study? Although details can be found later in L217-224 & L260-281, by slightly addressing which specific social response parameter was used in here would be better.

L137: Please describe a little bit more on the traps. Using other's established methods and citing them is okay, but do not just point the readers to find the reference themselves.

L139-140: When reporting geographic coordinate, latitude should go first, then longitude, so should be written as (xxxN xxxE).

L146, 149, 154, 181, 182, 198, 202, 203, 214, 220: it would be easier to read and follow if the reporting style and units are unified. For example, authors used both 1.5 m and 150 cm, another is x hour and x h. For the dimensions, authors could consider: "xx cm length * xx cm width * xx cm height" in full or so, but not just indicate the height. Furthermore, please go through all the spacing between the number and unit, as some are missing (for example, L145: 30 g, L146: xx cm, L158: 4 pm).

L168: (Table S1) is fine, delete "in Supplement".

L290: Table S4? Or should be Table S2 for PCA results?

L312-313: Repeated sentence.

L339: What is a “good” weather? Please define “good weather”.

L461: Please add references. L460-462 is an argument, but only one reference for one side.

Figure 1: For 1A, 1C and 1E, I think they are not being supportive for the topic: prior captive experience affects responses. 1B, 1D and 1F are okay, but compared to 1B and 1D, Figure S2 and S3 should be put in main text.

Figure S2 & S3: Should be in the main text (Materials and Methods) but not as supplementary. Please also label the perch number for the cage figure, especially number 2 and 5, as they were used as variable in this study.

Methods of PCA

Authors should indicate what variables have been included in the analysis in the main text, I can only find them in the supplementary materials. In here, L285-292 is more about the results rather then methods, these should be put in the results part instead.

Methods of MMM

L314-317: Need to rephase this part, which is confusing, and I cannot tell the fixed, random effects or covariates apart in one go. I have to go through other Tables and even the R script to figure it out. For easy reading, the authors could consider the use of numbering for each test, such as 1) model 1…, 2) model 2…etc, and state the variables, fixed, random effects and covariates included in the model.

2. Experimental Design

A. Sampling (Catching Location and Methods)

According to L137-142, the Great Tit were captured from two different locations: 1) Konnevesi Research Station (62.616°N 26.346°E) and 2) city of Jyväskylä (62.253°N 25.750°E).

I have two major concerns on this: 1) the two places are roughly 50 km (31 miles) apart; 2) they are two different types of habitats (rural and urban) from what I saw on Google Map. These suggest the Great Tit could have population level differences.

I understand the availability of sampling locations can be very limited with all sorts of factors, but why not choose only one of the locations in this case? Given the multiple dimensions of this study, using birds from distinctive populations would only add extra, unnecessary factor, which would affect the statistical power of the analysis. Or if it is aimed to study whether Great Tit from different populations have the same response, more than two sampling locations would be needed.

I would like to know the following:

- How many Great Tit was caught from each site?
- Both sites trap birds on every sampling day? Or just from one site each day?

B. Exploration Part

I realised the authors designed this experiment heavily based on Dingemanse *et al*. 2002, as well as a few other studies, which I found these studies are kind of a standard for the field of birds' exploration behaviour in novel environment. However, I have concerns regarding a few approaches, and just saying it is a standard alone cannot justify the methods.

a. Exploration in Room

In L187-190 an L238-240, authors indicated the birds were given 10 minutes each to explore the room and record their behaviour for the first 2 minutes after they started exploring. If the bird only started exploring after 8 minutes, the trial continues for 2 minutes. I have the following questions:

1. Any time had been given to the birds to get familiar with the room? Just like for the exploration test in cage.
2. Why gave the birds 10 minutes but only count their behaviour for 2 minutes?
3. For the birds started exploring the room once the trial begins, did they get return to their individual housing manually after 2 minutes and stop the trial, or remain in the room until the time is up?
4. For those started after 8 minutes, I am going to use a very extreme example here, let's say started in the last 10 seconds. This bird would have been in the room for almost 12 minutes. As from my point of view, birds could start gathering information of the room when they were still in the individual black boxes. The front door was opened.
5. I have concerns that such approach is very likely to be biased towards getting a higher exploration score for some of the birds.

b. Exploration in Cage

L245-248, authors counted the birds' movement even if they went back to the acclimation side. Following are my concerns or questions:

1. How many birds were such case? L247 only indicated as "some birds". If this was common, the cage design or experimental methods need to be improved. This sounds a bit like the authors did not think of the birds would go back to the other side.
2. For the birds went back to the acclimation side, did they act similarly in both sides, or there were differences? The two sides were not with the same setting and dimensions, might cause different interpretation, and could have resulted in variation of exploration behaviour. Especially they had been given 5 minutes in the acclimation side beforehand,

that was not a “novel environment” anymore when compared to the exploration side.

3. For those birds stayed in exploration side, did the authors count in the score for the two perches and other variables in acclimation area? If not, the data collected is biased and inaccurate. Birds went back to acclimation side had a “larger” cage and more perches to explore than those did not.

C. Social Response Part

According to the example video for the mirror test, I would say the bird can see its mirrored image from the perch at the top-left, and the mirror is visible from all perches. The authors referred perch 2 as far from the mirror and out of view (given in Table S2), but I would consider the Great Tit can have social responses with the mirror, i.e. looking into the mirror.

Referring to the comments above, point 3 in the previous part, whether the perches on acclimation side also accounted for would affect this part, especially the authors calculated the scores according to the exploration trial in cage.

Those birds spent time in the acclimation area had less time in the exploration area. Hence, their chance to be on perch 1-5 were much less than those only stayed in exploration side. And during the mirror test, the acclimation side was not available to any birds. This obviously caused bias and inaccuracy for the data.

3. Validity of the Findings

A. Sampling or Data Treatment

Based on the nature of the two sampling locations (as mentioned in “Experimental Design” above), I would say the Great Tit could have population level differences in the very first place. However, I failed to find the authors have applied any control measures on this factor in their analysis.

I am surprised that the authors cited Dingemanse *et al*. (2012) paper, but did not consider the potential effect caused by the different in population level in this study, especially the two populations here were from different habitats. A recent piece, also from Dingemanse’s team, it is not directly addressing how exploration behaviour vary within and among populations, but worth a reading as well (Mouchet *et al*. 2021). Moreover, there are studies demonstrated birds from urban and rural area can be distinctive, including behavioural terms, personality, aggressiveness...etc. Some of those even studied Great Tit. For example, the study by Hardman & Dalesman (2018) worked on Great Tit, Thompson *et al*. (2018) was on Black-capped Chickadees. I put the references down below.

References:

1. Dingemanse *et al*. 2012. Variation in personality and behavioural plasticity across four populations of the great tit *Parus major*. *Journal of Animal Ecology*, 81, 116-126.
2. Hardman S.I. & Dalesman S. 2018. Repeatability and degree of territorial aggression differs among urban and rural great tits (*Parus major*). *Scientific Reports*, (2018) 8:5042. DOI: 10.1038/s41598-018-23463-7
3. Mouchet *et al*. 2021. Heterogeneous selection on exploration behavior within and among West European populations of a passerine bird. *PNAS*, 118(28), e2024994118. DOI: https://doi.org/10.1073/pnas.2024994118
4. Thompson *et al*. 2018. Urbanization and individual differences in exploration and plasticity. *Behavioural Ecology*, 29(6), 1415-1425. DOI: https://doi.org/10.1093/beheco/ary103

B. PCA

I'm a bit confused with the PCA methodology and results. According to Table S2, I found the PC1, PC2 and PC3 should be the "Difference in time spent of Perch 5 + Number of pecks at the mirror", "Time spent looking at the mirror" and "Difference in time spent of Perch 2 + Difference in number of movements" respectively, as their Eigenvector Scores are 0.653 & 0.608, -0.843 and 0.767 & 0.630 respectively. However, except PC2 is the same conclusion, I am on a different page for PC1 and PC3.

In L286-289, authors wrote that PC1 also taken the number of pecks and movements into account. On the other hand, L289-292 indicated the behaviour of PC2 and PC3 were originally included in PC1. If other parameters were also part of one other variable, then, what are the differences between them, and how to justify the use of PCA in the first place?

Furthermore, how about the other three perches? Why not included in the PCA?

C. MMM and Univariate

I am not familiar with the MCMCglmm function. Therefore, comments below may not be accurate or useful for this manuscript, especially on the R script.

OVERVIEW

The authors ran the MMM twice, because of the prior experience had been treated as scale (length of captivity) and nominal (yes/no). It is a nice touch to investigate both directions. However, I disagree to run them as separated tests. The prior experience (yes/no) is a fixed effect, and the length of captivity should be treated as covariate, a scale data cannot be put in as fixed, so only one model is required.

The fixed effects, random effects, and covariates are also very confusing. This may just because of the writing is not clear enough. Sex, age, test type and test order should be treated as fixed effects and not covariates, because they are nominal data.

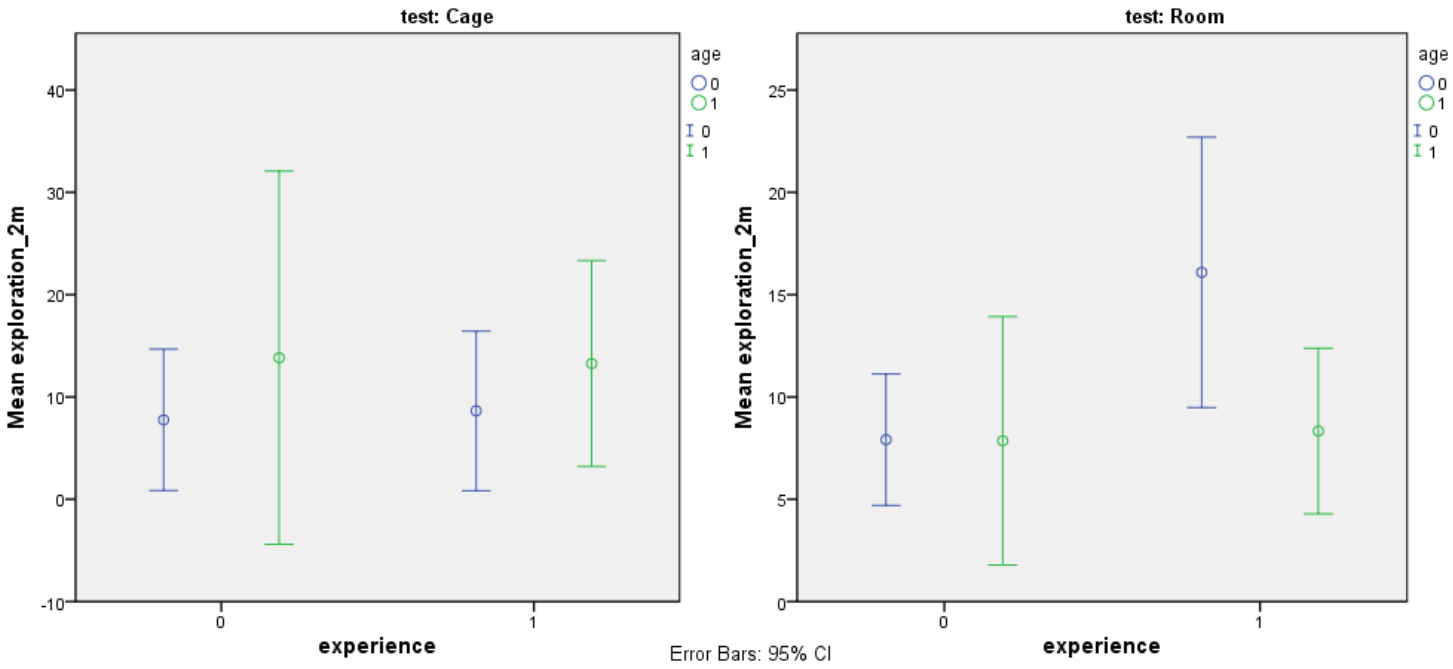
A univariate model was performed by the authors to investigate the interaction terms, which is an addition to the MMM. I cannot see why this has to be separated tests. The MMM should be able to handle interactions in one go.
I also have questions on other aspects, please find the details as follow:

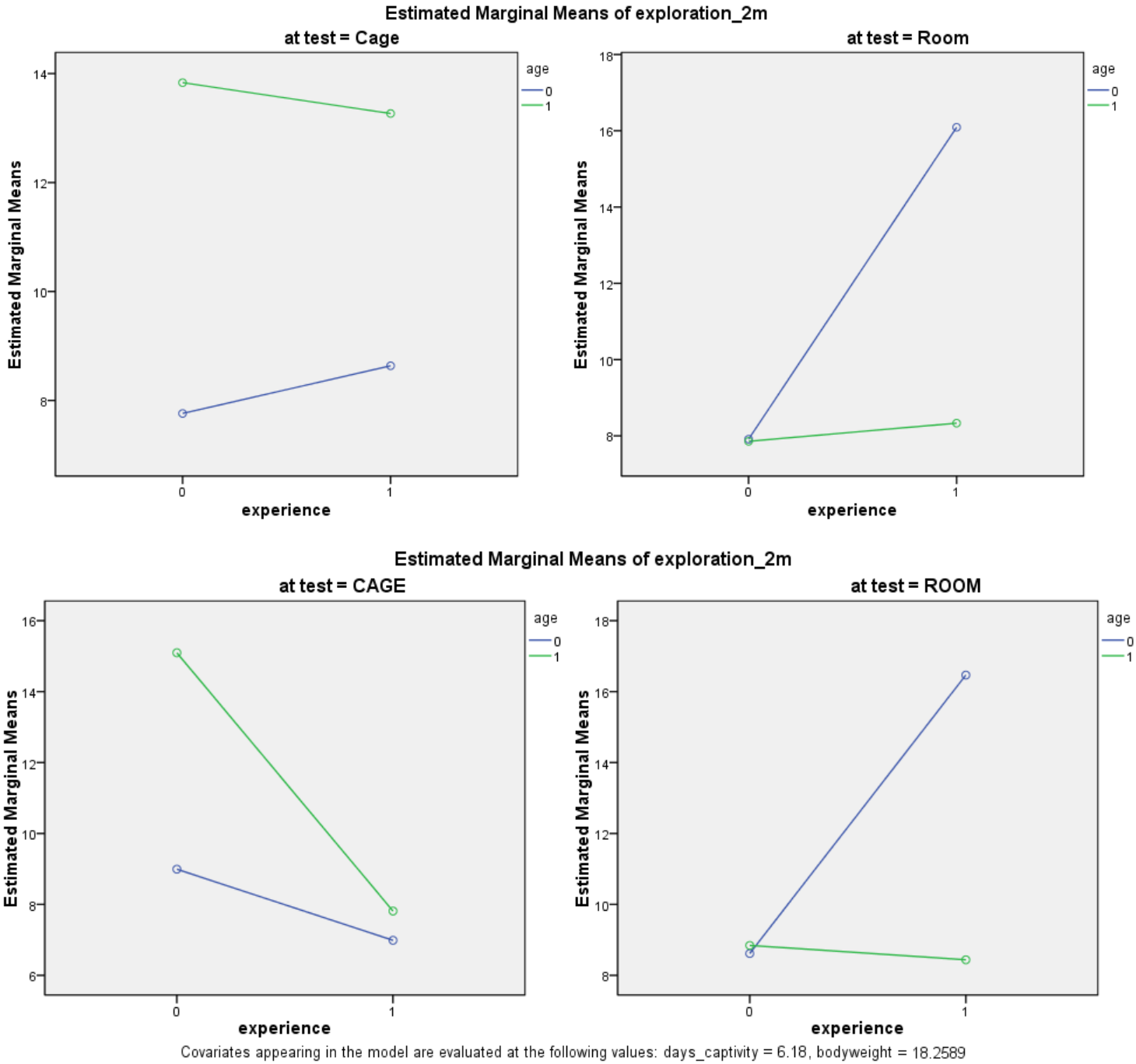
a. Exploration Data

i. Statistical Analysis

The physical environment of the room and cage were completely different, the settings, dimensions, even the temperature of the lighting were not the same. I see the room and cage as two different novel environments. The methods including handling was quite different between the two trials as well. From the test results provided by the authors, the test type (room or cage) did have effects on the exploration (Table 2 and S3, L368-370). This suggests the authors cannot assume birds have the same exploration behaviour and regardless the actual environment.
In the univariate models performed by the authors, the interaction terms with test type have not been investigated thoroughly, not all fixed effects combinations were included. I am not familiar with the analysis that the authors used, so I went for simple linear models to look at the interaction briefly. I suggest the authors to go through these interactions: 1) test type*age, 2) experience (nominal)*sex*test order, and 3) experience (nominal)*sex*test order*test.

The following figures: 1) Mean exploration score for both test types; 2) Plots from linear model without any covariates; and 3) Plots from linear model with the two covariates correction: days of captivity and bodyweight.





ii. Statements Based on Results

Authors concluded birds with prior captivity experience explore more, especially for first winter birds, they showed a stronger response than the adults also with prior experience.

Based on the figures above, I have the following comments:

1. The exploration behaviour was test type dependent in this study. Although whether they were significant or not would be depended on the analysis.
2. The authors' findings seem only fit for first winter birds in the room trial, but not in cage. Adults were more exploratory than first winter birds in the cage test, no matter those

adults had prior experience or not. All three figures showed the same.

3. Let's assume the statement is accurate first. The prior experience, or the carryover effect, does this only affect first winter birds and not for adults? What would happen if a first winter bird from this study was recapture a year later or so? Or I should ask was there any adult bird studied in this study, was captured during its first winter from the previous studies listed in Table S1? If so, what are their performance?

Based on all the above, I have doubt for the authors' conclusion. I understand there are lots of limitation and this is a preliminary work, so has to be built on assumptions, and many questions cannot be answered, but I think more can be done.

b. Table 2 & Table S3

I cannot tell the differences between the two tables. I noticed some of the figures are slightly different, but they are almost the same. From the methods mentioned by the authors, the MMM had been performed twice to cover the two types of captivity experience. This means each fixed effect and covariate has two results. I wonder which is which for the two Tables, and how the authors decided to use which output to be put in the Results in main text.

<u>Correlations</u>

As the nature of PC1 is questionable, any analysis and results involving PC1 should be reanalysed. For the exploration data, I suggest the authors to treat the data from the room and cage separately.

4. Additional Comments

I find the topic interesting and potentially important to the field. The idea that the authors trying to investigate here sounds solid in theory, and very likely to influence animal behaviour studies afterwards.

However, even though this is a preliminary work to investigate the idea, I have major concerns regarding multiple aspects of this study, from the study design, execution, analysis, to the findings and conclusions drawn.

I genuinely hope the comments above could help the authors to improve this work, and get this idea publish. Definitely looking forward to it.

A reject is recommended for this manuscript.