

## Social groups with diverse personalities mitigate physiological stress in a songbird

Csongor I. Vágási, Attila Fülöp, Gergely Osváth, Péter L. Pap, Janka Péntzes, Zoltán Benkő, Ádám Z. Lendvai and Zoltán Barta

### Article citation details

*Proc. R. Soc. B* **288**: 20203092.

<http://dx.doi.org/10.1098/rspb.2020.3092>

### Review timeline

Original submission: 20 May 2020  
1st revised submission: 11 December 2020  
2nd revised submission: 7 January 2021  
Final acceptance: 8 January 2021

Note: Reports are unedited and appear as submitted by the referee. The review history appears in chronological order.

## Review History

### RSPB-2020-1169.R0 (Original submission)

#### Review form: Reviewer 1

##### Recommendation

Major revision is needed (please make suggestions in comments)

**Scientific importance: Is the manuscript an original and important contribution to its field?**

Good

**General interest: Is the paper of sufficient general interest?**

Good

**Quality of the paper: Is the overall quality of the paper suitable?**

Good

**Is the length of the paper justified?**

Yes

**Should the paper be seen by a specialist statistical reviewer?**

No

**Do you have any concerns about statistical analyses in this paper? If so, please specify them explicitly in your report.**

Yes

**It is a condition of publication that authors make their supporting data, code and materials available - either as supplementary material or hosted in an external repository. Please rate, if applicable, the supporting data on the following criteria.**

**Is it accessible?**

Yes

**Is it clear?**

Yes

**Is it adequate?**

Yes

**Do you have any ethical concerns with this paper?**

No

### **Comments to the Author**

This manuscript focused on a very interesting question – that is whether the amount of within-group differences in the personality characteristics of its members affected individual physiological state. This is a broad subject area that touches a number of disciplines in the natural and social sciences, but as the authors indicate, no experimental study in wild animals has assembled groups to make diverse and uniform groups and assessed its impact on biomarkers of health and condition. Nicely done.

They captured house sparrows, assessed their exploratory behavior using a well-established behavioral assay and then grouped them into four treatments: random, low exploratory, high exploratory, variable. There were 4 replicates of each treatment and each treatment contained 10 total birds. They took pre-treatment blood sample 9 days after capture and measured body size/mass and then housed the birds for another 9 days in their treatment. They then took a post-treatment blood sample at the end of the treatment period and measured body size/mass. They measured body condition and used the blood samples to quantify some standard biomarkers of health.

There were no differences in any of the indices of condition or health before putting the birds in the treatment groups. The results for post-treatment blood samples are very interesting but require some careful explanation. Specifically, the authors show that their “variable” treatment group (where individuals with low and high exploration were mixed) did have higher variance in exploration compared to the low, high, or random groups (Fig. 1B), when exploration diversity was estimated using the Shannon index, the diversity was highest in the “random” treatment group. They then find that birds in the “random treatment” group with the highest Shannon diversity in personality increased in body condition and had a decrease in the heterophil-to-lymphocyte (H/L) ratio (suggestive of “less stress”) and some evidence of a reduction in a biomarker of oxidative stress (MDA) but not changes in two other measures of innate immunity.

They then take a different approach where they average the group level change in body condition, H/L ratio, and MDA for each of 4 replicates of each treatment group and regress this upon the Shannon diversity of exploration within the group. They find that groups with higher diversity in personality (at least for exploration) had an increase in body condition, reduction in H/L ratio and MDA.

This is fascinating as it shows experimentally that individuals can benefit from a physiological/condition perspective by joining groups that are more diverse in exploration personality. Why that is the case is uncertain but the authors present result from other studies in the Discussion as possible mechanisms.

I really liked reading this manuscript and thought it was well-written and a broad subject of general interest. I do have some suggestions for the authors.

1) I find the two analyses that can give slightly different results to be a bit confusing and in need of greater explanation. What I mean is that you have two ways of analyzing the same dataset (Fig. 1 and Fig. 2). I think you want to better justify the inclusion of the Shannon diversity analyses as they come across as being post-hoc. Also, why not have the individual changes in this analysis? You could have Shannon diversity per group, within-individual change in the measures, and random effects for individual (as you now have pre- and post-treatment samples) and group (as you would have multiple replicates per group)? If you are to do the analyses in Figure 2, I wouldn't see why you wouldn't want to increase the power here.

2) I find the within-individual changes (response variables in Figure 1) and average within-individual changes (Figure 2) to be at odds with other more suitable approaches. Why wouldn't you use the pre-treatment and post-treatment value for each individual (2 samples per individual) and then have treatment and your other fixed effects in here along with individual identity as a random effect? This is similar to my comment above but could expand your power. Also, I don't see why the data are presented as they are in Fig. 1 and Fig. 2 when they should be reaction norm figures so that readers can interpret the results better. I really advocate for this approach and this would be very helpful for readers to see the data.

3) I think more balance needs to be given to the biomarkers of health used here. While some studies suggest that simple markers like H/L or MDA might be indicative of health, others do not. This just needs more nuance here and acknowledgement of the negative results from other studies.

4) Some more details are required here such as how the variable groups were assembled. It seems random but then there was a random treatment. Did the authors expect that diversity index of the random group would be higher than the variable treatment group? If not, they should acknowledge this and the post-hoc nature of this analysis (if that is the case). I also think some more details about the house sparrows at this site would be helpful. For example, could you indicate the typical size of flocks in this species where you caught them? Is 10 usual? How long does it take the groups to form? It seems like 9 days is not that long for group formation. Do they form in any time of the year? Some of this may be in the supplementary materials but I think some issues (like the variable treatment group) should be described in the main text.

5) The authors nicely discuss some of the possible ways in which higher diversity within a group could benefit individual group members in the Discussion. It would be nice if they could speculate on what the mechanism is in their study - how does enhanced personality within the group affect individuals in this environment where food is not limited? I realize some of this is in lines 218-219 but what do you mean by type and intensity of interactions?

6) This manuscript cites three papers by Pruitt and each paper that is cited needs to be carefully vetted by the authors of this manuscript if they are to cite them. I cannot vouch for pubpeer but there are ways to see who collected the data in those papers in which Pruitt is a co-author or senior author. For example, the

Other minor comments:

Line 40 - think this should be "variation within groups"

Line 86 – I think introductory material to the species and why this species is crucial to be in the Introduction

Line 102 – I think this should be “a pre-treatment blood sample”

Line 106 – I’d like to know more here about this – was it 5 low and 5 high exploratory birds? Tell us more about the variable treatment.

Line 158 – why were both the dependent and predictor variables z-transformed here but only the predictor variables were z-transformed above (line 137-139)

Line 152 – the rationale for using the group average change rather than the individual change needs to be better explained and introduced. It does seem like doing the same analysis twice but more explanation could aid my understanding. It seems like some of this in the Results (lines 177-181)?

I think Figure 1 should be a reaction norm figure.

If the data were collected over multiple years, was year included as a fixed effect?

## Review form: Reviewer 2

### **Recommendation**

Major revision is needed (please make suggestions in comments)

**Scientific importance: Is the manuscript an original and important contribution to its field?**

Good

**General interest: Is the paper of sufficient general interest?**

Good

**Quality of the paper: Is the overall quality of the paper suitable?**

Marginal

**Is the length of the paper justified?**

Yes

**Should the paper be seen by a specialist statistical reviewer?**

No

**Do you have any concerns about statistical analyses in this paper? If so, please specify them explicitly in your report.**

No

**It is a condition of publication that authors make their supporting data, code and materials available - either as supplementary material or hosted in an external repository. Please rate, if applicable, the supporting data on the following criteria.**

**Is it accessible?**

No

**Is it clear?**

N/A

**Is it adequate?**

N/A

**Do you have any ethical concerns with this paper?**

No

**Comments to the Author**

RSPB-2020-1169

Social groups with diverse personalities mitigate physiological stress in a songbird

The authors present experimental data on how individuals respond to more or less diverse social environments, with diversity being focussed on personality score. They formed 24 flocks of house sparrows, which differed in the composition of their members' personality: some groups were more homogenous than others (i.e. high and low diversity of personality). The authors found that the sparrows' body condition improved, and two out of four measures of physiological and oxidative stress decreased with increasing group-level diversity of personality. While the study aims to address an interesting gap in the animal behavioural ecology literature, I am not convinced that all interpretation of the results is sufficiently supported by the current data.

My main points of concern are:

(1)

The authors only measure short-term changes in body condition and so-called physiological stress response, but seem to claim evolutionary consequences (e.g., L 233, "Our results provide a physiological mechanism that could also be responsible for the evolutionary maintenance of behavioural diversity in social groups"). Such statements in the discussion and elsewhere should be adjusted (e.g. "living in social groups with diverse composition can provide [short-term] benefits in terms of reduced physiological stress"), and the authors should carefully discuss the evolutionary relevance of the observed changes.

Also claims about potential health benefits in humans, from belonging to a diversely composed team, might be a bit premature and should be carefully revised and potentially backed-up by additional literature.

(2)

The introduction in its current form is not presenting testable hypotheses, founded on knowledge from previous research. For example, there are no references to why individuals are expected to differ consistently in their behaviour in the first place, or why naturally occurring groups may differ in their members' personality composition. Only then can one ask how different group compositions are expected to influence individual group members. I think the paragraphs LL 203-216 and LL 225ff (from the discussion) should be moved to the introduction.

Further, the authors added a few references to studies on human behaviour in the introduction, which might generally be acceptable, but it is confusing when the references suggest that the current study addresses gaps in the literature on human behaviour. For example, in line 65 ("but human studies with experimental manipulation of group composition and actual health measurements are still lacking") hints towards this gap being addressed in the current paper.

(3)

Figure Fig. 1A-C and the first sentence of the results section should not be presented as one of the authors' apparent main results: "The groups in the four experimental treatments differed in the mean, variance and Shannon diversity index of their personality composition" - this is not surprising given that groups were formed based on individuals' personality scores. It's good to see the data as proof of concept, but it should not be presented as a main result.

Other comments:

I am not familiar with the assessment of physiological parameter to describe individuals' condition, but can the authors elaborate on whether a single pre- and a single post-treatment measure is sufficiently accurate, and have other studies used a similar design? For example, when assessing cort metabolites, it is common practice to average multiple samples per individual.

L 169, "The differences in physiological condition among treatment groups (Fig. 1D-F) are mostly congruent with the group differences in personality diversity (cf. Fig. 1C) rather than with the mean or variance of personality of the groups (cf. Fig. 1A-B)." - How did the authors assess similarity?

L 180, "we computed the group-level mean of within-individual change in each physiological trait by averaging the change values of the 10 birds per group" - this approach potentially removes (i.e. averages out) important inter-individual differences in responses. It's surprising the authors chose this approach given the focus of the study. Can the authors present data on inter-individual variability in change in physiological traits? I doubt that all individuals were responding in a similar direction and magnitude.

The authors include interaction terms in their models but don't interpret and discuss significant interactions in their models. From the introduction I understood that the interaction between individual personality and group personality composition was a main point of interest to the authors.

Minor comments:

L 39: I don't agree that "social groups were once considered aggregates of similar individuals"; for example, differences in group members' behaviour or morphology have long been acknowledged in primate groups or groups of large social carnivores

L 47, "Social groups can largely differ in their personality composition": Not clear if this refers to individuals within groups differing in their personality, that some groups are more homogenous than other groups, or that on average some groups consist of more pro-active than reactive individuals. Please specify.

L 59: replace "performance" with "(social) behaviour" or another more specific term describing the context

L 165: remove "clearly"

## Decision letter (RSPB-2020-1169.R0)

08-Jul-2020

Dear Dr Vágási:

I am writing to inform you that your manuscript RSPB-2020-1169 entitled "Social groups with diverse personalities mitigate physiological stress in a songbird" has, in its current form, been rejected for publication in Proceedings B.

This action has been taken on the advice of referees, who have recommended that substantial revisions are necessary. With this in mind we would be happy to consider a resubmission, provided the comments of the referees are fully addressed. However please note that this is not a provisional acceptance.

The resubmission will be treated as a new manuscript. However, we will approach the same reviewers if they are available and it is deemed appropriate to do so by the Editor. Please note that resubmissions must be submitted within six months of the date of this email. In exceptional circumstances, extensions may be possible if agreed with the Editorial Office. Manuscripts submitted after this date will be automatically rejected.

Please find below the comments made by the referees, not including confidential reports to the Editor, which I hope you will find useful. If you do choose to resubmit your manuscript, please upload the following:

- 1) A 'response to referees' document including details of how you have responded to the comments, and the adjustments you have made.
- 2) A clean copy of the manuscript and one with 'tracked changes' indicating your 'response to referees' comments document.
- 3) Line numbers in your main document.

To upload a resubmitted manuscript, log into <http://mc.manuscriptcentral.com/prsb> and enter your Author Centre, where you will find your manuscript title listed under "Manuscripts with Decisions." Under "Actions," click on "Create a Resubmission." Please be sure to indicate in your cover letter that it is a resubmission, and supply the previous reference number.

Sincerely,  
Dr Maurine Neiman  
mailto:proceedingsb@royalsociety.org

Associate Editor

Comments to Author:

Two reviewers have now commented on the manuscript 'Social groups with diverse personalities mitigate physiological stress in a songbird'. Whereas both found the general topic of the study interesting, they identified problems with the study design (set-up of experimental groups), criticised the analysis approach, and questioned the robustness of the findings. They also thought that several of the conclusions were too bold given the evidence presented. All of these issues will need to be thoroughly and effectively addressed in a resubmission.

Reviewer(s)' Comments to Author:

Referee: 1

Comments to the Author(s)

This manuscript focused on a very interesting question – that is whether the amount of within-group differences in the personality characteristics of its members affected individual physiological state. This is a broad subject area that touches a number of disciplines in the natural and social sciences, but as the authors indicate, no experimental study in wild animals has assembled groups to make diverse and uniform groups and assessed its impact on biomarkers of health and condition. Nicely done.

They captured house sparrows, assessed their exploratory behavior using a well-established behavioral assay and then grouped them into four treatments: random, low exploratory, high exploratory, variable. There were 4 replicates of each treatment and each treatment contained 10 total birds. They took pre-treatment blood sample 9 days after capture and measured body size/mass and then housed the birds for another 9 days in their treatment. They then took a post-treatment blood sample at the end of the treatment period and measured body size/mass. They measured body condition and used the blood samples to quantify some standard biomarkers of health.

There were no differences in any of the indices of condition or health before putting the birds in the treatment groups. The results for post-treatment blood samples are very interesting but require some careful explanation. Specifically, the authors show that their “variable” treatment group (where individuals with low and high exploration were mixed) did have higher variance in exploration compared to the low, high, or random groups (Fig. 1B), when exploration diversity was estimated using the Shannon index, the diversity was highest in the “random” treatment group. They then find that birds in the “random treatment” group with the highest Shannon diversity in personality increased in body condition and had a decrease in the heterophil-to-lymphocyte (H/L) ratio (suggestive of “less stress”) and some evidence of a reduction in a biomarker of oxidative stress (MDA) but not changes in two other measures of innate immunity.

They then take a different approach where they average the group level change in body condition, H/L ratio, and MDA for each of 4 replicates of each treatment group and regress this upon the Shannon diversity of exploration within the group. They find that groups with higher diversity in personality (at least for exploration) had an increase in body condition, reduction in H/L ratio and MDA.

This is fascinating as it shows experimentally that individuals can benefit from a physiological/condition perspective by joining groups that are more diverse in exploration personality. Why that is the case is uncertain but the authors present result from other studies in the Discussion as possible mechanisms.

I really liked reading this manuscript and thought it was well-written and a broad subject of general interest. I do have some suggestions for the authors.

1) I find the two analyses that can give slightly different results to be a bit confusing and in need of greater explanation. What I mean is that you have two ways of analyzing the same dataset (Fig. 1 and Fig. 2). I think you want to better justify the inclusion of the Shannon diversity analyses as they come across as being post-hoc. Also, why not have the individual changes in this analysis? You could have Shannon diversity per group, within-individual change in the measures, and random effects for individual (as you now have pre- and post-treatment samples) and group (as you would have multiple replicates per group)? If you are to do the analyses in Figure 2, I wouldn't see why you wouldn't want to increase the power here.

2) I find the within-individual changes (response variables in Figure 1) and average within-individual changes (Figure 2) to be at odds with other more suitable approaches. Why wouldn't you use the pre-treatment and post-treatment value for each individual (2 samples per individual) and then have treatment and your other fixed effects in here along with individual identity as a random effect? This is similar to my comment above but could expand your power. Also, I don't see why the data are presented as they are in Fig. 1 and Fig. 2 when they should be reaction norm figures so that readers can interpret the results better. I really advocate for this approach and this would be very helpful for readers to see the data.

3) I think more balance needs to be given to the biomarkers of health used here. While some studies suggest that simple markers like H/L or MDA might be indicative of health, others do not. This just needs more nuance here and acknowledgement of the negative results from other studies.

4) Some more details are required here such as how the variable groups were assembled. It seems random but then there was a random treatment. Did the authors expect that diversity index of the random group would be higher than the variable treatment group? If not, they should acknowledge this and the post-hoc nature of this analysis (if that is the case). I also think some more details about the house sparrows at this site would be helpful. For example, could you indicate the typical size of flocks in this species where you caught them? Is 10 usual? How long does it take the groups to form? It seems like 9 days is not that long for group formation. Do they



form in any time of the year? Some of this may be in the supplementary materials but I think some issues (like the variable treatment group) should be described in the main text.

5) The authors nicely discuss some of the possible ways in which higher diversity within a group could benefit individual group members in the Discussion. It would be nice if they could speculate on what the mechanism is in their study – how does enhanced personality within the group affect individuals in this environment where food is not limited? I realize some of this is in lines 218-219 but what do you mean by type and intensity of interactions?

6) This manuscript cites three papers by Pruitt and each paper that is cited needs to be carefully vetted by the authors of this manuscript if they are to cite them. I cannot vouch for pubpeer but there are ways to see who collected the data in those papers in which Pruitt is a co-author or senior author. For example, the

Other minor comments:

Line 40 – think this should be “variation within groups”

Line 86 – I think introductory material to the species and why this species is crucial to be in the Introduction

Line 102 – I think this should be “a pre-treatment blood sample”

Line 106 – I’d like to know more here about this – was it 5 low and 5 high exploratory birds? Tell us more about the variable treatment.

Line 158 – why were both the dependent and predictor variables z-transformed here but only the predictor variables were z-transformed above (line 137-139)

Line 152 – the rationale for using the group average change rather than the individual change needs to be better explained and introduced. It does seem like doing the same analysis twice but more explanation could aid my understanding. It seems like some of this in the Results (lines 177-181)?

I think Figure 1 should be a reaction norm figure.

If the data were collected over multiple years, was year included as a fixed effect?

Referee: 2

Comments to the Author(s)

RSPB-2020-1169

Social groups with diverse personalities mitigate physiological stress in a songbird

The authors present experimental data on how individuals respond to more or less diverse social environments, with diversity being focussed on personality score. They formed 24 flocks of house sparrows, which differed in the composition of their members’ personality: some groups were more homogenous than others (i.e. high and low diversity of personality). The authors found that the sparrows’ body condition improved, and two out of four measures of physiological and oxidative stress decreased with increasing group-level diversity of personality. While the study aims to address an interesting gap in the animal behavioural ecology literature, I am not convinced that all interpretation of the results is sufficiently supported by the current data.

My main points of concern are:

(1)

The authors only measure short-term changes in body condition and so-called physiological stress response, but seem to claim evolutionary consequences (e.g., L 233, “Our results provide a physiological mechanism that could also be responsible for the evolutionary maintenance of behavioural diversity in social groups”). Such statements in the discussion and elsewhere should be adjusted (e.g. “living in social groups with diverse composition can provide [short-term] benefits in terms of reduced physiological stress”), and the authors should carefully discuss the evolutionary relevance of the observed changes.

Also claims about potential health benefits in humans, from belonging to a diversely composed team, might be a bit premature and should be carefully revised and potentially backed-up by additional literature.

(2)

The introduction in it's current form is not presenting testable hypotheses, founded on knowledge from previous research. For example, there are no references to why individuals are expected to differ consistently in their behaviour in the first place, or why naturally occurring groups may differ in their members' personality composition. Only then can one ask how different group compositions are expected to influence individual group members. I think the paragraphs LL 203-216 and LL 225ff (from the discussion) should be moved to the introduction.

Further, the authors added a few references to studies on human behaviour in the introduction, which might generally be acceptable, but it is confusing when the references suggest that the current study addresses gaps in the literature on human behaviour. For example, in line 65 (“but human studies with experimental manipulation of group composition and actual health measurements are still lacking”) hints towards this gap being addressed in the current paper.

(3)

Figure Fig. 1A-C and the first sentence of the results section should not be presented as one of the authors' apparent main results: “The groups in the four experimental treatments differed in the mean, variance and Shannon diversity index of their personality composition” – this is not surprising given that groups were formed based on individuals' personality scores. It's good to see the data as proof of concept, but it's should not be presented as a main result.

Other comments:

I am not familiar with the assessment of physiological parameter to describe individuals' condition, but can the authors elaborate on whether a single pre- and a single post-treatment measure is sufficiently accurate, and have other studies used a similar design? For example, when assessing cort metabolites, it is common practice to average multiple samples per individual.

L 169, “The differences in physiological condition among treatment groups (Fig. 1D-F) are mostly congruent with the group differences in personality diversity (cf. Fig. 1C) rather than with the mean or variance of personality of the groups (cf. Fig. 1A-B).” – How did the authors assess similarity?

L 180, “we computed the group-level mean of within-individual change in each physiological trait by averaging the change values of the 10 birds per group” – this approach potentially removes (i.e. averages out) important inter-individual differences in responses. It's surprising the authors chose this approach given the focus of the study. Can the authors present data on inter-individual variability in change in physiological traits? I doubt that all individuals were responding in a similar direction and magnitude.

The authors include interaction terms in their models but don't interpret and discuss significant interactions in their models. From the introduction I understood that the interaction between individual personality and group personality composition was a main point of interest to the authors.

Minor comments:

L 39: I don't agree that "social groups were once considered aggregates of similar individuals"; for example, differences in group members' behaviour or morphology have long been acknowledged in primate groups or groups of large social carnivores

L 47, "Social groups can largely differ in their personality composition": Not clear if this refers to individuals within groups differing in their personality, that some groups are more homogenous than other groups, or that on average some groups consist of more pro-active than reactive individuals. Please specify.

L 59: replace "performance" with "(social) behaviour" or another more specific term describing the context

L 165: remove "clearly"

## Author's Response to Decision Letter for (RSPB-2020-1169.R0)

See Appendix A.

## RSPB-2020-3092.R0

### Review form: Reviewer 1

#### **Recommendation**

Accept with minor revision (please list in comments)

**Scientific importance: Is the manuscript an original and important contribution to its field?**

Good

**General interest: Is the paper of sufficient general interest?**

Good

**Quality of the paper: Is the overall quality of the paper suitable?**

Good

**Is the length of the paper justified?**

Yes

**Should the paper be seen by a specialist statistical reviewer?**

No

**Do you have any concerns about statistical analyses in this paper? If so, please specify them explicitly in your report.**

Yes

**It is a condition of publication that authors make their supporting data, code and materials available - either as supplementary material or hosted in an external repository. Please rate, if applicable, the supporting data on the following criteria.**

**Is it accessible?**

Yes

**Is it clear?**

No

**Is it adequate?**

Yes

**Do you have any ethical concerns with this paper?**

No

### **Comments to the Author**

This is the second time I have reviewed this manuscript. I want to commend and thank the authors for responding to all of my queries in such a comprehensive and logical way. This really helps with reviewing manuscripts and I appreciate the time and effort they put in here.

In re-reading the manuscript and their responses to the reviewer comments, I am once again encouraged by this study. I find the many changes they made to be satisfactory but a few issues come up in this revision:

- 1) The changes to statistical analyses I had suggested all seem appropriate although they strengthen or weaken some of the effects.
- 2) Including the reaction norm figures in the supplementary results would be useful here given the interest in understanding why some individuals vary in their responses to the social group treatments.
- 3) If you are using GLMMs with binomial errors, you would need to describe the dispersion parameters to ensure that they are not overdispersed? I believe these are binary variables (lines 190-191) so might say they didn't need to be reported here.
- 4) I do like the revised Figure 1 but encourage the authors to 1) make it colour-blind friendly and 2) put the raw (individual) data points on the figure as well. I also think the x-axis needs a label? I'd encourage them to put a line legend as an inset on Figure 2 as well. This would really help readers.
- 5) I thank the authors for the description regarding the papers authored by Pruitt that were cited here. Sounds like they vetted them and it is up to their discretion to cite them. It is unfortunate that a brief sentence couldn't describe this process given that other readers will not be aware this process occurred.

## **Review form: Reviewer 2**

### **Recommendation**

Accept as is

**Scientific importance: Is the manuscript an original and important contribution to its field?**

Good

**General interest: Is the paper of sufficient general interest?**

Excellent

**Quality of the paper: Is the overall quality of the paper suitable?**

Good

**Is the length of the paper justified?**

Yes

**Should the paper be seen by a specialist statistical reviewer?**

No

**Do you have any concerns about statistical analyses in this paper? If so, please specify them explicitly in your report.**

No

**It is a condition of publication that authors make their supporting data, code and materials available - either as supplementary material or hosted in an external repository. Please rate, if applicable, the supporting data on the following criteria.**

**Is it accessible?**

Yes

**Is it clear?**

Yes

**Is it adequate?**

Yes

**Do you have any ethical concerns with this paper?**

No

#### **Comments to the Author**

I read the revised manuscript on how flocks that vary in the composition of their group members' personality scores differ in their members' physiological parameters. The authors have addressed all my previously made (minor) suggestions and replied to my comments in a sufficient matter, by providing additional explanations in both the manuscript, the ESM, and in the authors' response letter.

I only have very minor comments: One suggestion that may help improve the interpretability of figure 2: I would suggest introducing some scatter to the plot, because the open symbols are partly overlaid by the closed symbols. Also, maybe pairing open symbols with dashed lines and closed symbols with solid lines might be more intuitive than the current pairing.

I saw the authors provided the raw data to their manuscript, which is well-prepared with annotations and explanations, and I would encourage the authors to also provide the script used for the data analysis, to further add to the transparency of the data analysis.

## **Decision letter (RSPB-2020-3092.R0)**

31-Dec-2020

Dear Dr Vágási

I am pleased to inform you that your manuscript RSPB-2020-3092 entitled "Social groups with diverse personalities mitigate physiological stress in a songbird" has been accepted for publication in Proceedings B.

The referee(s) have recommended publication, but also suggest some minor revisions to your manuscript. Therefore, I invite you to respond to the referee(s)' comments and revise your manuscript. Because the schedule for publication is very tight, it is a condition of publication that you submit the revised version of your manuscript within 7 days. If you do not think you will be able to meet this date please let us know.

To revise your manuscript, log into <https://mc.manuscriptcentral.com/prsb> and enter your Author Centre, where you will find your manuscript title listed under "Manuscripts with Decisions." Under "Actions," click on "Create a Revision." Your manuscript number has been appended to denote a revision. You will be unable to make your revisions on the originally submitted version of the manuscript. Instead, revise your manuscript and upload a new version through your Author Centre.

When submitting your revised manuscript, you will be able to respond to the comments made by the referee(s) and upload a file "Response to Referees". You can use this to document any changes you make to the original manuscript. We require a copy of the manuscript with revisions made since the previous version marked as 'tracked changes' to be included in the 'response to referees' document.

Before uploading your revised files please make sure that you have:

- 1) A text file of the manuscript (doc, txt, rtf or tex), including the references, tables (including captions) and figure captions. Please remove any tracked changes from the text before submission. PDF files are not an accepted format for the "Main Document".
- 2) A separate electronic file of each figure (tiff, EPS or print-quality PDF preferred). The format should be produced directly from original creation package, or original software format. PowerPoint files are not accepted.
- 3) Electronic supplementary material: this should be contained in a separate file and where possible, all ESM should be combined into a single file. All supplementary materials accompanying an accepted article will be treated as in their final form. They will be published alongside the paper on the journal website and posted on the online figshare repository. Files on figshare will be made available approximately one week before the accompanying article so that the supplementary material can be attributed a unique DOI.

Online supplementary material will also carry the title and description provided during submission, so please ensure these are accurate and informative. Note that the Royal Society will not edit or typeset supplementary material and it will be hosted as provided. Please ensure that the supplementary material includes the paper details (authors, title, journal name, article DOI). Your article DOI will be 10.1098/rspb.[paper ID in form xxxx.xxxx e.g. 10.1098/rspb.2016.0049].

- 4) A media summary: a short non-technical summary (up to 100 words) of the key findings/importance of your manuscript.

- 5) Data accessibility section and data citation

It is a condition of publication that data supporting your paper are made available either in the electronic supplementary material or through an appropriate repository (<https://royalsociety.org/journals/authors/author-guidelines/#data>).

In order to ensure effective and robust dissemination and appropriate credit to authors the dataset(s) used should be fully cited. To ensure archived data are available to readers, authors should include a 'data accessibility' section immediately after the acknowledgements section. This should list the database and accession number for all data from the article that has been made publicly available, for instance:

- DNA sequences: Genbank accessions F234391-F234402
- Phylogenetic data: TreeBASE accession number S9123
- Final DNA sequence assembly uploaded as online supplemental material
- Climate data and MaxEnt input files: Dryad doi:10.5521/dryad.12311

NB. From April 1 2013, peer reviewed articles based on research funded wholly or partly by RCUK must include, if applicable, a statement on how the underlying research materials – such as data, samples or models – can be accessed. This statement should be included in the data accessibility section.

If you wish to submit your data to Dryad (<http://datadryad.org/>) and have not already done so you can submit your data via this link

[http://datadryad.org/submit?journalID=RSPB&manu=\(Document not available\)](http://datadryad.org/submit?journalID=RSPB&manu=(Document%20not%20available)) which will take you to your unique entry in the Dryad repository. If you have already submitted your data to dryad you can make any necessary revisions to your dataset by following the above link. Please see <https://royalsociety.org/journals/ethics-policies/data-sharing-mining/> for more details.

6) For more information on our Licence to Publish, Open Access, Cover images and Media summaries, please visit <https://royalsociety.org/journals/authors/author-guidelines/>.

Once again, thank you for submitting your manuscript to Proceedings B and I look forward to receiving your revision. If you have any questions at all, please do not hesitate to get in touch.

Sincerely,

Dr Maurine Neiman

mailto:proceedingsb@royalsociety.org

Associate Editor

Board Member

Comments to Author:

The authors have thoroughly revised the manuscript and most of the reviewers' previous comments are now addressed. There are a number of issues remaining, which the authors should be able to address in a further revision:

1. Please address the reviewers' comments.
2. l. 161: MDA is a measure of oxidative damage, not oxidative stress (these terms should not be used interchangeably). Please change throughout the manuscript.
3. l. 177: explain what a high activity of the humoral innate immune system means. Are individuals with high levels very immunocompetent or very susceptible to infection (i.e. infected)?
4. l. 178-179: present detailed results in the ESM. Also present analyses of the correlations among the different physiological traits in the ESM.
5. Please state dispersion parameters of GLMMs with binomial errors (l. 211).
6. Present the interaction effects before the main effects in the results section. It is typically not meaningful to interpret main effects if the interaction effect is significant.
7. l. 294 'variety is delighting' is unclear. Please reword.
8. Discussion: The results of the variable group seem to contradict some of the main conclusions. Discuss why there were no physiological benefits in the variable group. This group is diverse as well but effects on SMI and H/L are very different in the variable and random groups. Include a section in the Discussion where you discuss differences between the variable and random groups explicitly, and why such differences might occur.
9. l. 309 – 317: this section is repetitive, revise.

### Reviewer(s)' Comments to Author:

Referee: 1

Comments to the Author(s).

This is the second time I have reviewed this manuscript. I want to commend and thank the authors for responding to all of my queries in such a comprehensive and logical way. This really helps with reviewing manuscripts and I appreciate the time and effort they put in here.

In re-reading the manuscript and their responses to the reviewer comments, I am once again encouraged by this study. I find the many changes they made to be satisfactory but a few issues come up in this revision:

- 1) The changes to statistical analyses I had suggested all seem appropriate although they strengthen or weaken some of the effects.
- 2) Including the reaction norm figures in the supplementary results would be useful here given the interest in understanding why some individuals vary in their responses to the social group treatments.
- 3) If you are using GLMMs with binomial errors, you would need to describe the dispersion parameters to ensure that they are not overdispersed? I believe these are binary variables (lines 190-191) so might say they didn't need to be reported here.
- 4) I do like the revised Figure 1 but encourage the authors to 1) make it colour-blind friendly and 2) put the raw (individual) data points on the figure as well. I also think the x-axis needs a label? I'd encourage them to put a line legend as an inset on Figure 2 as well. This would really help readers.
- 5) I thank the authors for the description regarding the papers authored by Pruitt that were cited here. Sounds like they vetted them and it is up to their discretion to cite them. It is unfortunate that a brief sentence couldn't describe this process given that other readers will not be aware this process occurred.

Referee: 2

Comments to the Author(s).

I read the revised manuscript on how flocks that vary in the composition of their group members' personality scores differ in their members' physiological parameters. The authors have addressed all my previously made (minor) suggestions and replied to my comments in a sufficient matter, by providing additional explanations in both the manuscript, the ESM, and in the authors' response letter.

I only have very minor comments: One suggestion that may help improve the interpretability of figure 2: I would suggest introducing some scatter to the plot, because the open symbols are partly overlaid by the closed symbols. Also, maybe pairing open symbols with dashed lines and closed symbols with solid lines might be more intuitive than the current pairing.

I saw the authors provided the raw data to their manuscript, which is well-prepared with annotations and explanations, and I would encourage the authors to also provide the script used for the data analysis, to further add to the transparency of the data analysis.

### Author's Response to Decision Letter for (RSPB-2020-3092.R0)

See Appendix B.



## Decision letter (RSPB-2020-3092.R1)

08-Jan-2021

Dear Dr Vágási

I am pleased to inform you that your manuscript entitled "Social groups with diverse personalities mitigate physiological stress in a songbird" has been accepted for publication in Proceedings B.

You can expect to receive a proof of your article from our Production office in due course, please check your spam filter if you do not receive it. PLEASE NOTE: you will be given the exact page length of your paper which may be different from the estimation from Editorial and you may be asked to reduce your paper if it goes over the 10 page limit.

If you are likely to be away from e-mail contact please let us know. Due to rapid publication and an extremely tight schedule, if comments are not received, we may publish the paper as it stands.

If you have any queries regarding the production of your final article or the publication date please contact [procb\\_proofs@royalsociety.org](mailto:procb_proofs@royalsociety.org)

Your article has been estimated as being 9 pages long. Our Production Office will be able to confirm the exact length at proof stage.

### Open Access

You are invited to opt for Open Access, making your freely available to all as soon as it is ready for publication under a CCBY licence. Our article processing charge for Open Access is £1700.

Corresponding authors from member institutions

(<http://royalsocietypublishing.org/site/librarians/allmembers.xhtml>) receive a 25% discount to these charges. For more information please visit <http://royalsocietypublishing.org/open-access>.

### Paper charges

An e-mail request for payment of any related charges will be sent out shortly. The preferred payment method is by credit card; however, other payment options are available.

### Electronic supplementary material:

All supplementary materials accompanying an accepted article will be treated as in their final form. They will be published alongside the paper on the journal website and posted on the online figshare repository. Files on figshare will be made available approximately one week before the accompanying article so that the supplementary material can be attributed a unique DOI.

You are allowed to post any version of your manuscript on a personal website, repository or preprint server. However, the work remains under media embargo and you should not discuss it with the press until the date of publication. Please visit <https://royalsociety.org/journals/ethics-policies/media-embargo> for more information.

Thank you for your fine contribution. On behalf of the Editors of the Proceedings B, we look forward to your continued contributions to the Journal.

Sincerely,

Proceedings B

<mailto:proceedingsb@royalsociety.org>

# Appendix A

Editorial Board  
Proceedings B  
The Royal Society

11-Dec-2020

Resubmission: MS Reference Number RSPB-2020-1169

Dear Prof Barrett – Editor-in-Chief,  
Dear Dr Neiman – Handling Editor,

Please find enclosed our revised manuscript entitled "Social groups with diverse personalities mitigate physiological stress in a songbird." Our original manuscript (RSPB-2020-1169) was submitted to *Proceedings B* and was rejected after full peer review with the possibility of resubmission. We are grateful for the constructive comments raised by the Reviewers and for their positive tone. These comments definitely improved our manuscript. We thank their valuable help in the Acknowledgements (see lines #405–406).

We substantially revised our manuscript following the recommendations of the Reviewers. Here, we provide our detailed responses to each comment separately. Our responses are marked in boldface. We provide the line numbers where the modifications can be found in the revised manuscript with tracked changes. Additionally, we submit a clean copy of the revised manuscript.

We hope our revision and responses will be found appropriate. Thank you for your editorial work.

Sincerely yours,

Csongor Vágási, on behalf of all authors

Associate Editor

Two reviewers have now commented on the manuscript 'Social groups with diverse personalities mitigate physiological stress in a songbird'. Whereas both found the general topic of the study interesting, they identified problems with the study design (set-up of experimental groups), criticised the analysis approach, and questioned the robustness of the findings. They also thought that several of the conclusions were too bold given the evidence presented. All of these issues will need to be thoroughly and effectively addressed in a resubmission.

**We are happy to hear that both Reviewers found our study interesting and that our study fits the aim of Proceedings B. We are grateful for their constructive comments and thank their help in the Acknowledgements (see lines #405–406). We believe that we successfully addressed all their comments and adjusted the manuscript accordingly. Importantly, we changed the statistical analyses as suggested by Reviewer #1 and these changes did not modify the main results and conclusions. Finally, we toned down the evolutionary implications of our results as suggested by Reviewer #2.**

**We hope the Editors and Reviewers will find our revision appropriate.**

## Thank you for your Editorial work.

Referee: 1

This manuscript focused on a very interesting question – that is whether the amount of within-group differences in the personality characteristics of its members affected individual physiological state. This is a broad subject area that touches a number of disciplines in the natural and social sciences, but as the authors indicate, no experimental study in wild animals has assembled groups to make diverse and uniform groups and assessed its impact on biomarkers of health and condition. Nicely done.

They captured house sparrows, assessed their exploratory behavior using a well-established behavioral assay and then grouped them into four treatments: random, low exploratory, high exploratory, variable. There were 4 replicates of each treatment and each treatment contained 10 total birds. They took pre-treatment blood sample 9 days after capture and measured body size/mass and then housed the birds for another 9 days in their treatment. They then took a post-treatment blood sample at the end of the treatment period and measured body size/mass. They measured body condition and used the blood samples to quantify some standard biomarkers of health.

There were no differences in any of the indices of condition or health before putting the birds in the treatment groups. The results for post-treatment blood samples are very interesting but require some careful explanation. Specifically, the authors show that their “variable” treatment group (where individuals with low and high exploration were mixed) did have higher variance in exploration compared to the low, high, or random groups (Fig. 1B), when exploration diversity was estimated using the Shannon index, the diversity was highest in the “random” treatment group. They then find that birds in the “random treatment” group with the highest Shannon diversity in personality increased in body condition and had a decrease in the heterophil-to-lymphocyte (H/L) ratio (suggestive of “less stress”) and some evidence of a reduction in a biomarker of oxidative stress (MDA) but not changes in two other measures of innate immunity.

They then take a different approach where they average the group level change in body condition, H/L ratio, and MDA for each of 4 replicates of each treatment group and regress this upon the Shannon diversity of exploration within the group. They find that groups with higher diversity in personality (at least for exploration) had an increase in body condition, reduction in H/L ratio and MDA.

This is fascinating as it shows experimentally that individuals can benefit from a physiological/condition perspective by joining groups that are more diverse in exploration personality. Why that is the case is uncertain but the authors present result from other studies in the Discussion as possible mechanisms.

I really liked reading this manuscript and thought it was well-written and a broad subject of general interest. I do have some suggestions for the authors.

**We are glad that the Reviewer appreciated our work and gave a number of constructive comments. We addressed all of them in the revised manuscript. Importantly, we now use repeated-measures analyses as suggested by the Reviewer to make use of individual data of both pre- and post-treatment samples for both sets of analyses (i.e. treatment group differences on one hand and association of physiological responses with Shannon diversity on the other hand). This new analytical approach yielded results that are qualitatively similar to those in the original manuscript; although the results related to the H/L ratio, that were only weakly significant, became marginally significant (i.e.  $p < 0.1$ ). This repeated-measures approach, thus, largely supports our previous results and make the conclusions more conservative.**

1) I find the two analyses that can give slightly different results to be a bit confusing and in need of greater explanation. What I mean is that you have two ways of analyzing the same dataset (Fig. 1 and Fig. 2). I think you want to better justify the inclusion of the Shannon diversity analyses as they come across as being

post-hoc. Also, why not have the individual changes in this analysis? You could have Shannon diversity per group, within-individual change in the measures, and random effects for individual (as you now have pre- and post-treatment samples) and group (as you would have multiple replicates per group)? If you are to do the analyses in Figure 2, I wouldn't see why you wouldn't want to increase the power here.

**Thank you for this suggestion. Because this first concern is largely similar to the second one, we respond to them together after the second concern (see below).**

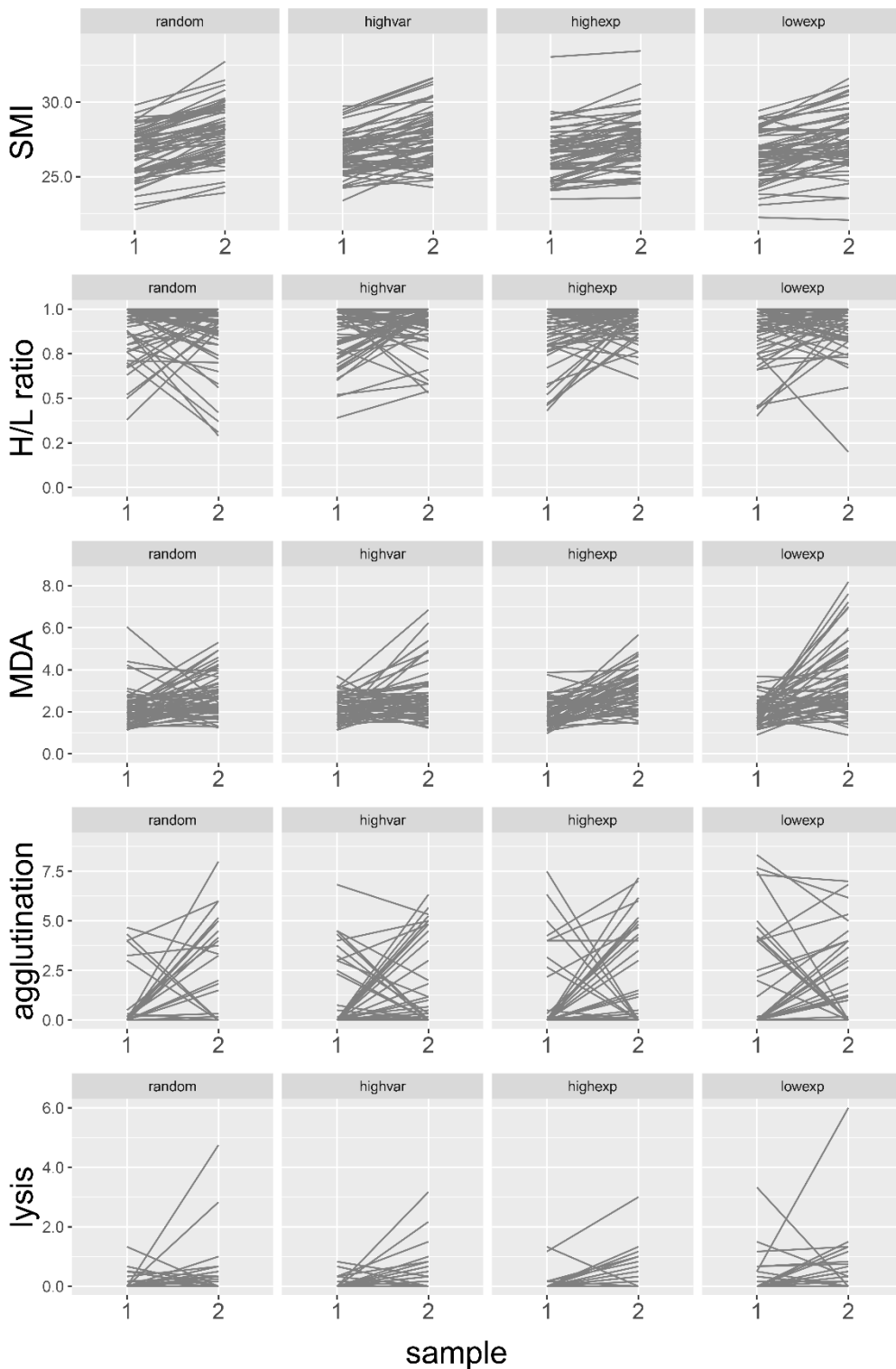
2) I find the within-individual changes (response variables in Figure 1) and average within-individual changes (Figure 2) to be at odds with other more suitable approaches. Why wouldn't you use the pre-treatment and post-treatment value for each individual (2 samples per individual) and then have treatment and your other fixed effects in here along with individual identity as a random effect? This is similar to my comment above but could expand your power. Also, I don't see why the data are presented as they are in Fig. 1 and Fig. 2 when they should be reaction norm figures so that readers can interpret the results better. I really advocate for this approach and this would be very helpful for readers to see the data.

**Thank you for these useful suggestions. We agree that the individual data of both sampling times and the corresponding repeated-measures analyses are more appropriate. Below we respond to all the suggestions and questions raised by the Reviewer under her/his first and second concern. The order of our responses follow the structure of the original manuscript instead of the order of Reviewers' suggestions and questions.**

The first set of analyses were meant to present the differences among the four treatment groups in terms of mean, variance and Shannon diversity of group personality composition (original figure 1a–c) or in terms of individual physiological response to treatment (original figure 1d–h). We reran the analyses using a repeated-measures approach. For this, we added to the data set a variable called “sampling event” with two levels for the two samples collected from the same individual (“pre-treatment” and “post-treatment”). We added “sampling event” as a fixed factor to the models and its interaction with “treatment” as well, besides the rest of the predictor variables presented in the original manuscript, and added individuals' ID as an additional random term (see Statistical procedures in the revised manuscript; lines #188–239). We found similar results to those presented in the original manuscript (where we used within-individual change values for the physiological variables, i.e. post- minus pre-treatment values) with the exception that the slightly significant treatment effect on H/L ratio became slightly non-significant (see table 1 and figure 1 in the revised manuscript at lines #529–534 and #554–555). Please note that in this new analysis the sampling event × social treatment interaction is of main interest (in contrast with the original analysis in which the social treatment main effect was in focus). We replaced the results in table 1 with these new repeated-measures results (see lines #529–534).

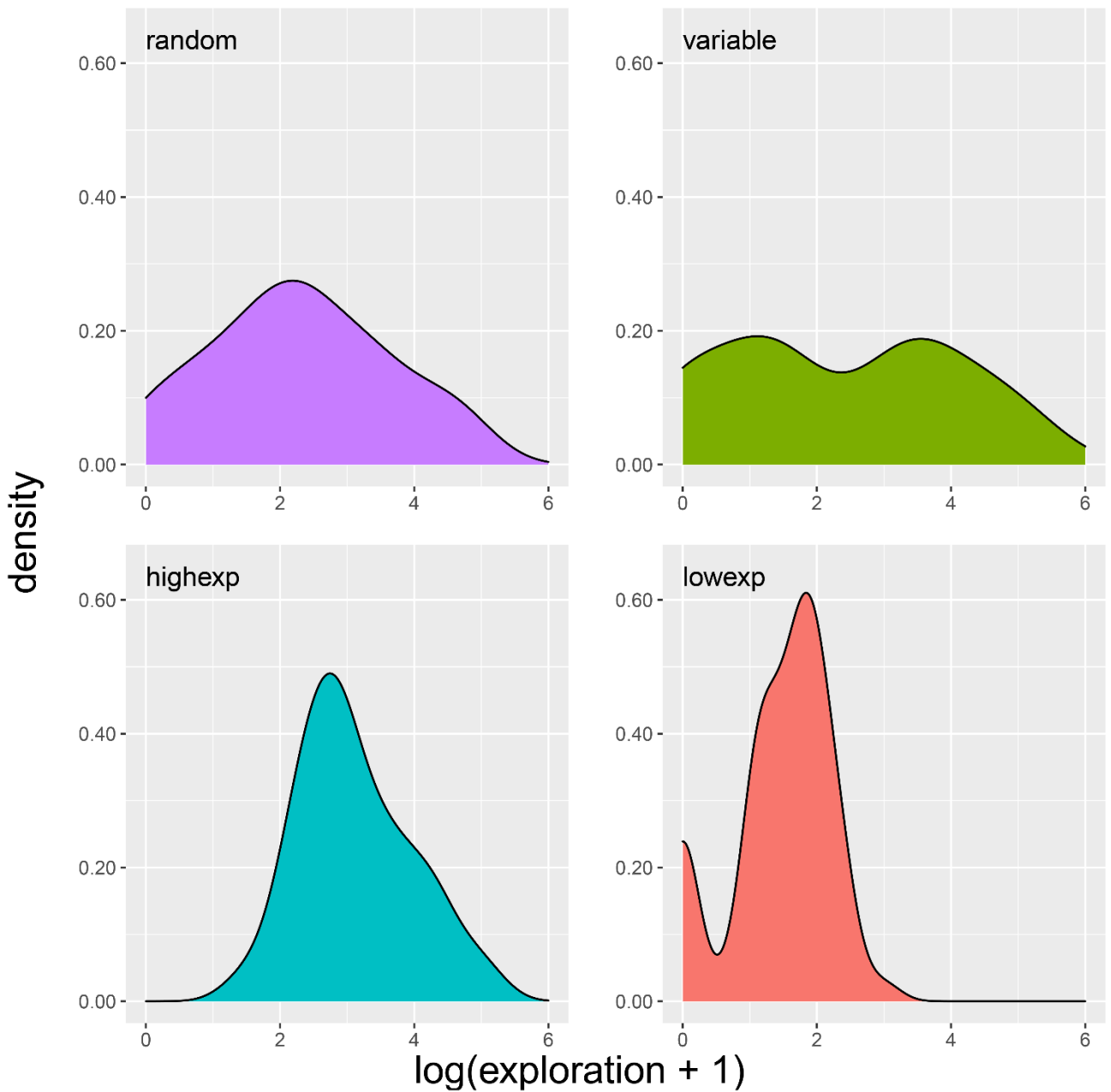
We also plotted the individual reaction norms for each of the five physiological response variables based on the raw data. However, we think that the individual reaction norm figures are difficult to read because within each treatment group there are 60 reaction norm lines (60 birds per treatment group). Below, we inserted these figures, but did not add them to either the revised manuscript or the Electronic Supplementary Material. Instead, we made a bar plot for each of the five response variables showing the model-predicted means and their SE for both the pre-treatment and post-treatment sampling events and we added these bar plots as figure 1 to the revised manuscript (see lines #554–555).

To summarize, we moved the original figure 1a–c panels into the Electronic Supplementary Material (see ESM figure S1a–c) to show that the experimental group assembly was successful, while the original figure 1d–h were changed to bar plots (see figure 1a–e in the revised manuscript at lines #554–555).



In accordance with the Reviewer's request, we now shortly justify the inclusion of Shannon diversity index in the Methods (see lines #141–160). Here in the responses, we explain it in more detail. The Reviewer is right that the Shannon diversity index came across *a posteriori*. The concept paper by Farine et al. (2015, Trends Ecol Evol), which is the conceptual foundation of our study, emphasizes that the mean and variance of a biological trait (in our case personality) within a social group are the primary measures of group phenotypic composition. Therefore, we developed our group assembly protocol *a priori* such to generate differences in mean and variance of personality (exploration score) among the four treatment groups (see ESM Methods), and the results met our expectations (see ESM figure S1a–b and related statistics in the revised manuscript at lines #141–160). Based on this concept, we expected that the differences among treatment groups in terms of individual physiological responses will mirror either the differences in mean personality of the group (i.e. low exploratory < random AND variable <

high exploratory or the reverse order; see figure S1a) or the differences in variance of personality within group (i.e. low AND high exploratory < random < variable or the reverse order; figure S1b). Note that for both mean and variance of group personality, the random group is intermediate. However, contrary to this expectation, we found that the random group was outstanding either alone or together with the variable group (see figure 1 and Results in the revised manuscript, and table S2 as well). This guided us to plot the distribution of exploration scores for each treatment group, which showed us that, while the random group has lower variance in personality than the variable group, their distributions are largely different. In the random group, the distribution is a wide, quasi-Gaussian, unimodal one, while it is wide but bimodal in the variable group (see distribution plots below that were also added to the Electronic Supplementary Material as figure S2). This is because the random group is a random sample of the entire range of exploration scores, while the variable group consists of five low- and five high-exploratory birds (see ESM Methods). We briefly explain these differences in distribution of the exploration scores in the revised manuscript as well (see lines #150–153).



The personality composition of the random group is therefore more diverse than that of the variable group. To capture this presumed difference, we computed the Shannon diversity for group personality composition and found that, indeed, the random group is composed of more diverse personalities than the variable group, and the latter is more diverse than the homogeneously composed low- and high-

exploratory groups (see ESM figure S1c and lines #156–160). This indicated to us, that potentially the diversity of personalities within the social group might explain the individual physiological responses to social treatment.

This hypothesis was tested by the second set of analyses in which we assessed whether group-level Shannon index is associated or not with the individual physiological responses to treatment. We found strong support for our hypothesis. Therefore, even if we blur the marked differences among treatment groups in personality composition and use this finer and more continuous scale of Shannon index, we arrive to the same main conclusion: “Social groups with diverse personalities mitigate physiological stress in a songbird.”

Now, in accordance with the Reviewer’s suggestion, we changed this second set of analyses as well by applying the repeated-measures procedure here too (see lines #225–230). Therefore, we added to the models the “sampling event” fixed factor and “sampling event × Shannon diversity” interaction as further predictors, while individual ID as an additional random effect. The results are qualitatively the same as those presented in the original manuscript with the exception that the weakly significant association between Shannon diversity and H/L ratio disappeared with this repeated-measures approach. Nevertheless, the main conclusion of the study remains. We now present these new statistics in the Results section (see lines #294–322 and the new table 2 at lines #535–540). We also changed the original figure 2, so that it presents the association between Shannon diversity and the individuals’ physiological state separately for the pre-treatment and post-treatment values (see the revised manuscript at lines #556–557). Note, however, that this revised figure 2 has only two panels showing SMI (condition) and MDA (oxidative damage), while the panel with H/L ratio was removed.

We hope that we answered all the questions raised by the Reviewer and that the ambiguous parts are clarified now.

3) I think more balance needs to be given to the biomarkers of health used here. While some studies suggest that simple markers like H/L or MDA might be indicative of health, others do not. This just needs more nuance here and acknowledgement of the negative results from other studies.

We are aware that the chosen physiological markers cannot describe in its entirety the highly complex physiological system of a vertebrate species. This also applies for any other marker we choose. The number of markers are always limited by the amount of blood sample, which is quite small in case of small-sized study organisms, such as the house sparrow.

Here we explain the rationale of choosing the physiological traits we used in this study. Changes in body condition take more time than changes in other physiological traits. Among the factors that affect body condition, exposure to stress stimuli is known to induce loss of body condition (Breuner et al. 2013 *Funct Ecol* 27: 24–36). Ultimately, impaired body condition can have widespread consequences for the organism. H/L ratio has been shown to correlate with the glucocorticoid stress response governed by hypothalamic-pituitary-adrenal axis (Davis et al. 2008 *Funct Ecol* 22: 760–772; Goessling et al. 2015 *Funct Ecol* 29: 1189–1196). H/L ratio is less sensitive to handling stress as compared with plasma levels of corticosterone (the main glucocorticoid in birds). MDA is a versatile marker of oxidative stress for two reasons. First, it indicates the damage to vital cell membrane lipids, the damage of which has substantial adverse consequences from the cell level to the organism level. Second, besides being a direct measure of oxidative damage, MDA is a pro-oxidant itself with a long half-life and hence can reach places far from its origin damaging other vital macromolecules (Del Rio et al. 2005 *Nutr Metab Cardiovasc Dis* 15: 316–328). Agglutination and lysis capacity of the plasma describes the activity of the humoral innate immune system and hence is an indicator of the first line of defence of vertebrate hosts against invading microorganisms (Matson et al. 2005 *Dev Comp Immunol* 29: 275–286). We added these short arguments to the Methods of the revised manuscript to highlight the relevance of choosing these physiological variables. See lines #167–181.

4) Some more details are required here such as how the variable groups were assembled. It seems random but then there was a random treatment. Did the authors expect that diversity index of the random group would be higher than the variable treatment group? If not, they should acknowledge this and the post-hoc nature of this analysis (if that is the case). I also think some more details about the house sparrows at this site would be helpful. For example, could you indicate the typical size of flocks in this species where you caught them? Is 10 usual? How long does it take the groups to form? It seems like 9 days is not that long for group formation. Do they form in any time of the year? Some of this may be in the supplementary materials but I think some issues (like the variable treatment group) should be described in the main text.

**In the Methods of the Electronic Supplementary Material under the “(e) Social treatment” subheading, we provided a detailed description of the protocol of assembling the four treatment groups with different personality composition. We mentioned this at line #118–120 in the original manuscript (see lines #136–141 in the revised manuscript). The description of social group assignment is quite lengthy, and hence, we cannot move all of these details into the main text due to the very stringent word count limit of Proceedings B. Nevertheless, we now indicate in the main text that this protocol is based on stratification of the individuals according to their exploration scores (see line #137).**

As we explained above, it is intuitively expected that the personality diversity is significantly higher in the random group as compared with the variable group. Therefore, computing the Shannon diversity index came across *a posteriori* and our results are suggestive that among-group difference in Shannon diversity (rather than mean or variance of personality) is important in explaining the individual physiological responses to social treatment.

**Regarding details on the social life of house sparrows in general and at our study site in particular, we added a new “Study species” subheading to the Methods of the revised manuscript with the following text: “The house sparrow (*Passer domesticus*) is arguably one of the most popular model organisms in animal ecology and evolution (Hanson et al. 2020 eLife 9: e52803). It is an ideal candidate to study social behaviour, since it exhibits a wide spectrum of social behaviour including colonial breeding, social foraging or communal roosting (Anderson TR 2006. *Biology of the ubiquitous house sparrow: from genes to populations*. Oxford University Press). At our study site, house sparrows are year-round residents, breed in cavities of stall buildings at the cattle farm and forage in flocks of various sizes, especially outside the breeding season. Flock sizes vary from small to medium, containing from a few birds up to some dozen individuals similarly to other populations (see Anderson 2006).” See lines #113–120.**

5) The authors nicely discuss some of the possible ways in which higher diversity within a group could benefit individual group members in the Discussion. It would be nice if they could speculate on what the mechanism is in their study – how does enhanced personality within the group affect individuals in this environment where food is not limited? I realize some of this is in lines 218–219 but what do you mean by type and intensity of interactions?

**We discussed the possible mechanisms a few lines above (see lines #203–216 in the original manuscript). We added another mechanism to the revised manuscript, which emphasizes the complementarity of different personalities based on a study in sticklebacks. Stickleback shoals solve better a two-stage food acquisition problem when the shoal contains fish that have experience with stage one and fish that have experience with stage two (termed “experience-pooling”; Webster et al. 2017, *Nat Ecol Evol* 1: 0135). Please note that we moved these sentences into the Introduction following the suggestion by Reviewer #2 (see lines #76–95 in the revised manuscript).**

6) This manuscript cites three papers by Pruitt and each paper that is cited needs to be carefully vetted by the authors of this manuscript if they are to cite them. I cannot vouch for pubpeer but there are ways to see who collected the data in those papers in which Pruitt is a co-author or senior author.



**We already did this assessment before the original manuscript was submitted to be sure that the papers signed by Jonathan Pruitt and cited in our manuscript have not been retracted or are not suspected to contain fabricated data. The papers left in the list of References are either concept/review papers that do not contain original data (Modlmeier et al. 2014 Anim Behav 89: 53–62) or primary research articles (Pruitt & Riechert 2011 Behav Ecol Sociobiol 65: 1055–1060 and Pruitt & Riechert 2009 Evolution 63: 2966–2973). Because there is no suspicion or concernment of misconduct in case of the latter two articles, which are based on original data, and thus have not been retracted yet, we kept these in the manuscript. However, we can remove them from the references if either the Reviewer or the Editor would like us to do so.**

Other minor comments:

Line 40 – think this should be “variation within groups”

**We corrected. See line #42.**

Line 86 – I think introductory material to the species and why this species is crucial to be in the Introduction

**We added a few sentences to the Methods about why the house sparrow is a good study organism to address the questions of our research. See lines #113–120.**

Line 102 – I think this should be “a pre-treatment blood sample”

**We corrected. See line #136.**

Line 106 – I’d like to know more here about this – was it 5 low and 5 high exploratory birds? Tell us more about the variable treatment.

**The word count limit set by Proceedings B does not permit to provide all the details of the methods. Therefore, we provided these details in the Methods of the Electronic Supplementary Material as stated at lines #118–120 in the original manuscript (see lines #136–141 in the revised manuscript). In brief, the variable group was assembled by choosing 5 low-exploratory and 5 high-exploratory birds, as the Reviewer correctly pointed out. Please, check the exact protocol in the Electronic Supplementary Material and our response above regarding the Shannon diversity index.**

Line 158 – why were both the dependent and predictor variables z-transformed here but only the predictor variables were z-transformed above (line 137-139)

**Probably our text was not clear enough here, but we Z-transformed both the dependent and continuous predictor variables in both analyses as we stated at lines #137–139 and lines #157–158 in the original manuscript. We restructured the “Statistical procedures” subheading of the Methods. Now we state clearly that all the continuous dependent variables and all continuous predictor variables were Z-transformed for both the first and second set of models (see lines #196–199 in the revised manuscript).**

Line 152 – the rationale for using the group average change rather than the individual change needs to be better explained and introduced. It does seem like doing the same analysis twice but more explanation could aid my understanding. It seems like some of this in the Results (lines 177-181)?

**We changed these analyses following the Reviewer's recommendations (see above). Therefore, this explanation is not relevant anymore, and the former explanation in the Results (lines #177–181 in the original manuscript) was deleted.**

I think Figure 1 should be a reaction norm figure.

**We made the plots suggested by the Reviewer. Please see the reaction norm figure and our response above. Note that we changed the original figure 1 to bar plots showing both the pre- and post-treatments values per each treatment group.**

If the data were collected over multiple years, was year included as a fixed effect?

**The Reviewer is right that the study was running over two years, but one winter season. Therefore, we do not think year should be included as a fixed effect especially because within season temporal effect was controlled for by including study replicate as random effect. The tables S2 and S3 provide the variance explained by this random effect.**

**We are grateful for the constructive comments of the Reviewer.**

Referee: 2

The authors present experimental data on how individuals respond to more or less diverse social environments, with diversity being focussed on personality score. They formed 24 flocks of house sparrows, which differed in the composition of their members' personality: some groups were more homogenous than others (i.e. high and low diversity of personality). The authors found that the sparrows' body condition improved, and two out of four measures of physiological and oxidative stress decreased with increasing group-level diversity of personality. While the study aims to address an interesting gap in the animal behavioural ecology literature, I am not convinced that all interpretation of the results is sufficiently supported by the current data.

**We are glad that the Reviewer appreciated our work and gave a number of constructive comments. We addressed all of them in the revised version of the manuscript.**

My main points of concern are:

(1) The authors only measure short-term changes in body condition and so-called physiological stress response, but seem to claim evolutionary consequences (e.g., L 233, "Our results provide a physiological mechanism that could also be responsible for the evolutionary maintenance of behavioural diversity in social groups"). Such statements in the discussion and elsewhere should be adjusted (e.g. "living in social groups with diverse composition can provide [short-term] benefits in terms of reduced physiological stress"), and the authors should carefully discuss the evolutionary relevance of the observed changes.

Also claims about potential health benefits in humans, from belonging to a diversely composed team, might be a bit premature and should be carefully revised and potentially backed-up by additional literature.

**In agreement with the Reviewer's suggestion, we reformulated our statements regarding the evolutionary implications of our results. Particularly, we toned down those related to the evolutionary maintenance of behavioural diversity and removed those about the evolution of leadership and**

**cooperation. See lines #30–32 and lines #368–370. We also removed claims about the health benefits in humans. See line #372 and lines #376–378.**

(2) The introduction in its current form is not presenting testable hypotheses, founded on knowledge from previous research. For example, there are no references to why individuals are expected to differ consistently in their behaviour in the first place, or why naturally occurring groups may differ in their members' personality composition. Only then can one ask how different group compositions are expected to influence individual group members. I think the paragraphs LL 203-216 and LL 225ff (from the discussion) should be moved to the introduction.

Further, the authors added a few references to studies on human behaviour in the introduction, which might generally be acceptable, but it is confusing when the references suggest that the current study addresses gaps in the literature on human behaviour. For example, in line 65 (“but human studies with experimental manipulation of group composition and actual health measurements are still lacking”) hints towards this gap being addressed in the current paper.

**We agree with the Reviewer that the Introduction should show why it is reasonable to ask how affiliation of similar or dissimilar personalities affect group members in social species. Therefore, we moved the Discussion sentences at lines #203–216 (original manuscript) to the Introduction, as the Reviewer suggested (see lines #76–95 in the revised manuscript). However, we kept the sentences from line #225 (original manuscript) in the Discussion because these are more of interpretation of the results by looking at their implications. See paragraph beginning at line #354 in the revised manuscript.**

(3) Figure Fig. 1A–C and the first sentence of the results section should not be presented as one of the authors' apparent main results: “The groups in the four experimental treatments differed in the mean, variance and Shannon diversity index of their personality composition” – this is not surprising given that groups were formed based on individuals' personality scores. It's good to see the data as proof of concept, but it's should not be presented as a main result.

**Indeed, we aimed with figure 1a–c to visualize the validity of our experimental protocol. Following the Reviewer's recommendation, we moved these results into the Methods of the revised manuscript (see lines #143–160) and the plots in the Electronic Supplementary Material (see ESM figure S1).**

Other comments:

I am not familiar with the assessment of physiological parameter to describe individuals' condition, but can the authors elaborate on whether a single pre- and a single post-treatment measure is sufficiently accurate, and have other studies used a similar design? For example, when assessing cort metabolites, it is common practice to average multiple samples per individual.

**In our previous studies, we found that there is high within-individual repeatability of the chosen physiological traits in house sparrows kept under similar conditions (see e.g. Pap et al. 2014 *Physiol Biochem Zool* 87: 729–739). This is indicative that an individual has a consistent physiological state that differs from other individuals' physiological state over time. Therefore, a single pre- and post-treatment sample is sufficiently accurate, and therefore this is a general protocol in physiological ecology studies using wild-living organisms.**

L 169, “The differences in physiological condition among treatment groups (Fig. 1D–F) are mostly congruent with the group differences in personality diversity (cf. Fig. 1C) rather than with the mean or variance of personality of the groups (cf. Fig. 1A–B).” – How did the authors assess similarity?

**This was a simple visual inspection. According to the request of Reviewer #1, we now explain in more detail the rationale of computing the Shannon diversity index (please see our response for Reviewer #1 above and the revised manuscript). Our supposition that the Shannon diversity index of group personality composition explains the group differences in physiological responses was formally tested in the second set of analyses, the results of which are presented at figure 2 and table 2 in the revised manuscript (see also table S3 in the ESM).**

L 180, “we computed the group-level mean of within-individual change in each physiological trait by averaging the change values of the 10 birds per group” – this approach potentially removes (i.e. averages out) important inter-individual differences in responses. It’s surprising the authors chose this approach given the focus of the study. Can the authors present data on inter-individual variability in change in physiological traits? I doubt that all individuals were responding in a similar direction and magnitude.

**The Reviewer is right that there is within-group variability both in direction and in magnitude of individual responses to treatment. The newer analyses are based on a repeated-measures procedure as requested by Reviewer #1, which grasps the among-individual variability both within and among groups. Please see above our responses given for Reviewer #1 and the newer analyses and results in the revised manuscript.**

The authors include interaction terms in their models but don’t interpret and discuss significant interactions in their models. From the introduction I understood that the interaction between individual personality and group personality composition was a main point of interest to the authors.

**After repeating the statistical analyses following the recommendations of Reviewer #1, the previously significant interactions between individual personality and social treatment have disappeared. Therefore, this issue is no longer relevant.**

Minor comments:

L 39: I don’t agree that “social groups were once considered aggregates of similar individuals”; for example, differences in group members’ behaviour or morphology have long been acknowledged in primate groups or groups of large social carnivores

**We accept the Reviewer's view and we modified the sentence accordingly. See lines #40–41.**

L 47, “Social groups can largely differ in their personality composition”: Not clear if this refers to individuals within groups differing in their personality, that some groups are more homogenous than other groups, or that on average some groups consist of more pro-active than reactive individuals. Please specify.

**Thank you for pointing this out. We reformulated this sentence, now it reads “Social groups can largely differ in their personality composition some being more homogenous, while others more heterogeneous.” See lines #49–50.**

L 59: replace “performance” with “(social) behaviour” or another more specific term describing the context

**We changed accordingly. See line #62.**

L 165: remove “clearly”

**We changed accordingly. See line #247.**

**We are grateful for the constructive comments of the Reviewer.**

## Appendix B

Editorial Board  
Proceedings B  
The Royal Society

07-Jan-2021

MS Reference Number RSPB-2020-3092

Dear Prof Barrett – Editor-in-Chief,

Dear Dr Neiman – Handling Editor,

Please find enclosed our revised manuscript entitled "Social groups with diverse personalities mitigate physiological stress in a songbird." We are glad that both Reviewers found our manuscript revision appropriate and our study suitable for Proceedings B. We appreciate again the Reviewers' constructive comments. We revised our manuscript following the recommendations of the Reviewers. Here, we provide our detailed responses to each comment separately. Our responses are marked in boldface. We provide the line numbers where the modifications can be found in the revised manuscript with tracked changes appended to this Response to Referees document. Additionally, we submit a clean copy of the revised manuscript.

We hope our revision and responses will be found appropriate. Thank you for your editorial work.

Sincerely yours,

Csongor Vágási, on behalf of all authors

Associate Editor

Board Member

The authors have thoroughly revised the manuscript and most of the reviewers' previous comments are now addressed. There are a number of issues remaining, which the authors should be able to address in a further revision:

**We are happy to hear that our manuscript revision was found satisfactory. We dealt with all the remaining issues in the revised manuscript:**

1. Please address the reviewers' comments.

**1. We responded to each comment raised by the two Reviewers. Please see our responses below.**

2. l. 161: MDA is a measure of oxidative damage, not oxidative stress (these terms should not be used interchangeably). Please change throughout the manuscript.

**2. We changed "oxidative stress" to "oxidative damage" throughout the manuscript, wherever we referred specifically to MDA. See lines #27, 63, 161, 244, 270, 274, 290 and 312.**

3. l. 177: explain what a high activity of the humoral innate immune system means. Are individuals with high levels very immunocompetent or very susceptible to infection (i.e. infected)?

**3. We specified that higher agglutination and lysis scores indicate a higher activity of the constitutive innate humoral immunity, i.e. birds with higher scores are more immune responsive (not more infected). See lines #177–178.**

4. l. 178-179: present detailed results in the ESM. Also present analyses of the correlations among the different physiological traits in the ESM.

**4. (i) We present in the Electronic Supplementary Material the full statistics separately for each of the five physiological response variables showing that treatment groups did not differ in the pre-treatment sample. We refer to these statistics in the revised main text. See line #180. (ii) We also present in the Electronic Supplementary Material the correlation between the five physiological variables for both the pre- and post-treatment samples (see table S2). The correlations are generally weak except for the positive association between agglutination and lysis scores. Therefore, these variables describe independent aspects of the physiological state. We added a reference to these correlations in the revised main text. See lines #180–182.**

5. Please state dispersion parameters of GLMMs with binomial errors (l. 211).

**5. We indeed used models with binomial error distribution to analyse agglutination and lysis data. Because these two response variables were binary (0 or 1), overdispersion for such models is not estimable.**

6. Present the interaction effects before the main effects in the results section. It is typically not meaningful to interpret main effects if the interaction effect is significant.

**6. We changed the order of presenting the results, moving the results of the sampling event × social treatment interaction in front of the main effects, as requested. See Results in the revised manuscript.**

7. l. 294 ‘variety is delighting’ is unclear. Please reword.

**7. We reworded as suggested; we now write “In contrast, our results are consonant with the old Latin proverb “Varietas delectat” [42]...” See lines #309–313.**

8. Discussion: The results of the variable group seem to contradict some of the main conclusions. Discuss why there were no physiological benefits in the variable group. This group is diverse as well but effects on SMI and H/L are very different in the variable and random groups. Include a section in the Discussion where you discuss differences between the variable and random groups explicitly, and why such differences might occur.

**8. Thank you for this suggestion. We added a new paragraph to the Discussion of the revised manuscript in which we explicitly deal with why the variable and random groups might differ. See lines #318–335.**

9. l. 309 – 317: this section is repetitive, revise.

**9. We revised this section of the Discussion. See lines #341–352.**

**We hope the Editors and Reviewers will find our revision appropriate. Thank you for your Editorial work.**

Referee: 1

This is the second time I have reviewed this manuscript. I want to commend and thank the authors for responding to all of my queries in such a comprehensive and logical way. This really helps with reviewing manuscripts and I appreciate the time and effort they put in here.

**We are glad that the Reviewer appreciated our revision and gave further constructive comments.**

In re-reading the manuscript and their responses to the reviewer comments, I am once again encouraged by this study. I find the many changes they made to be satisfactory but a few issues come up in this revision:

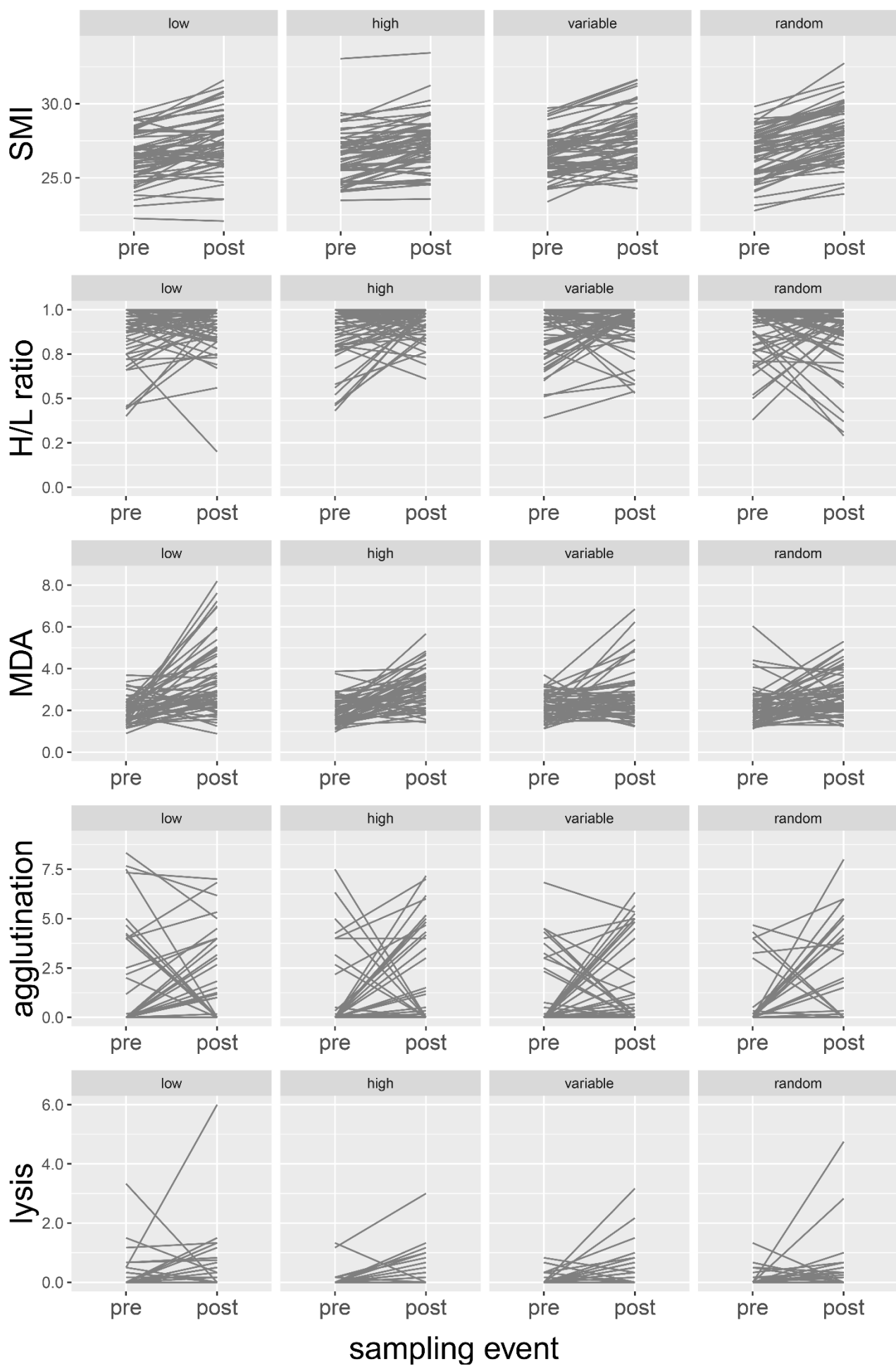
1) The changes to statistical analyses I had suggested all seem appropriate although they strengthen or weaken some of the effects.

**1. The Reviewer is right and we admitted that some results weakened by the new statistical procedure. Specifically, (i) the treatment effect on H/L ratio became marginally non-significant, while being weakly significant in the original analyses, and (ii) the association between Shannon diversity of personality and H/L ratio became non-significant, while being weakly significant in the original analysis. Despite these changes, the main conclusion of our study remains unchanged.**

2) Including the reaction norm figures in the supplementary results would be useful here given the interest in understanding why some individuals vary in their responses to the social group treatments.

**2. We added the individual reaction norm figure to the Electronic Supplementary Material as figure S3 and referred to this figure in the main text. See lines #246–247. Figure S3:**





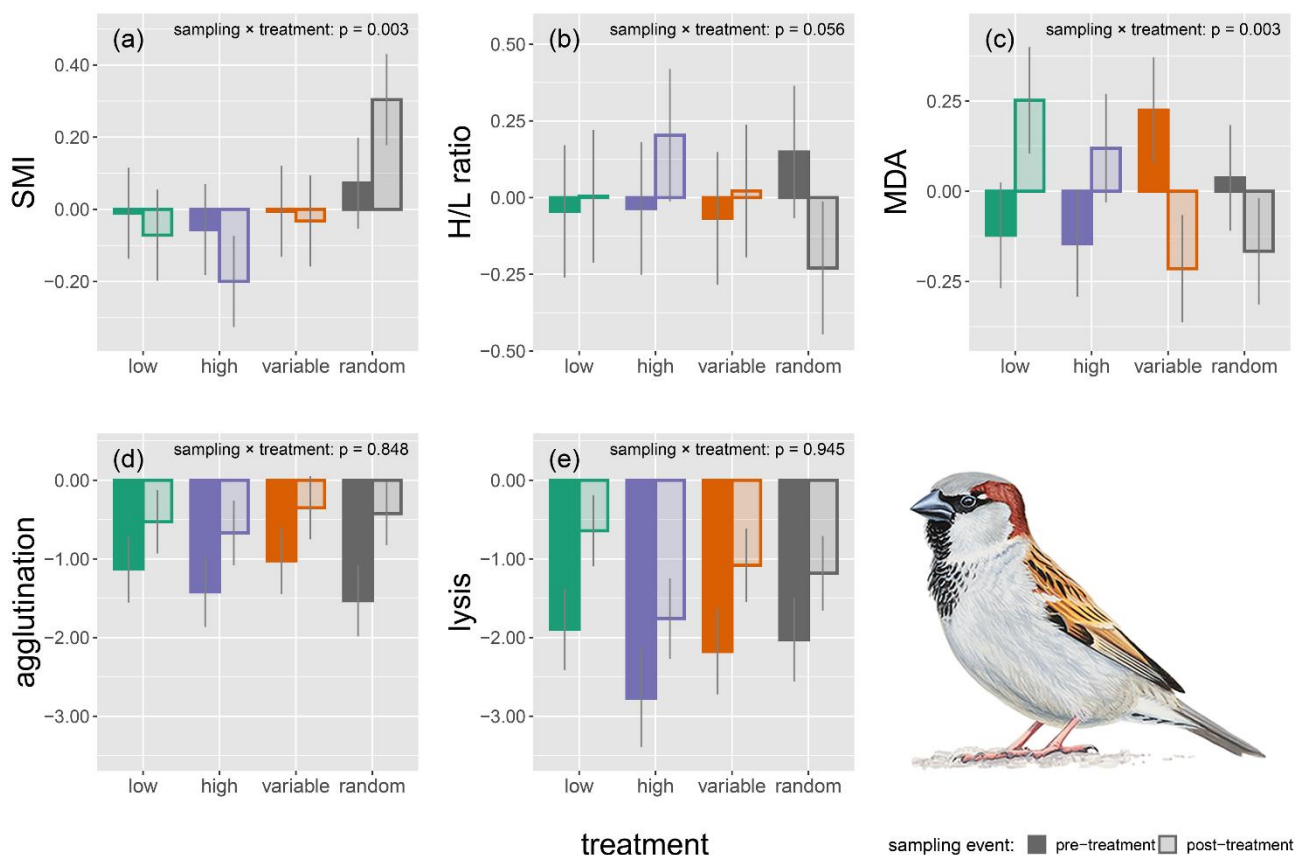
3) If you are using GLMMs with binomial errors, you would need to describe the dispersion parameters to ensure that they are not overdispersed? I believe these are binary variables (lines 190-191) so might say they didn't need to be reported here.

**3. We indeed used models with binomial error distribution to analyse agglutination and lysis data. Because these two response variables were binary (0 or 1), overdispersion for such models is not estimable.**

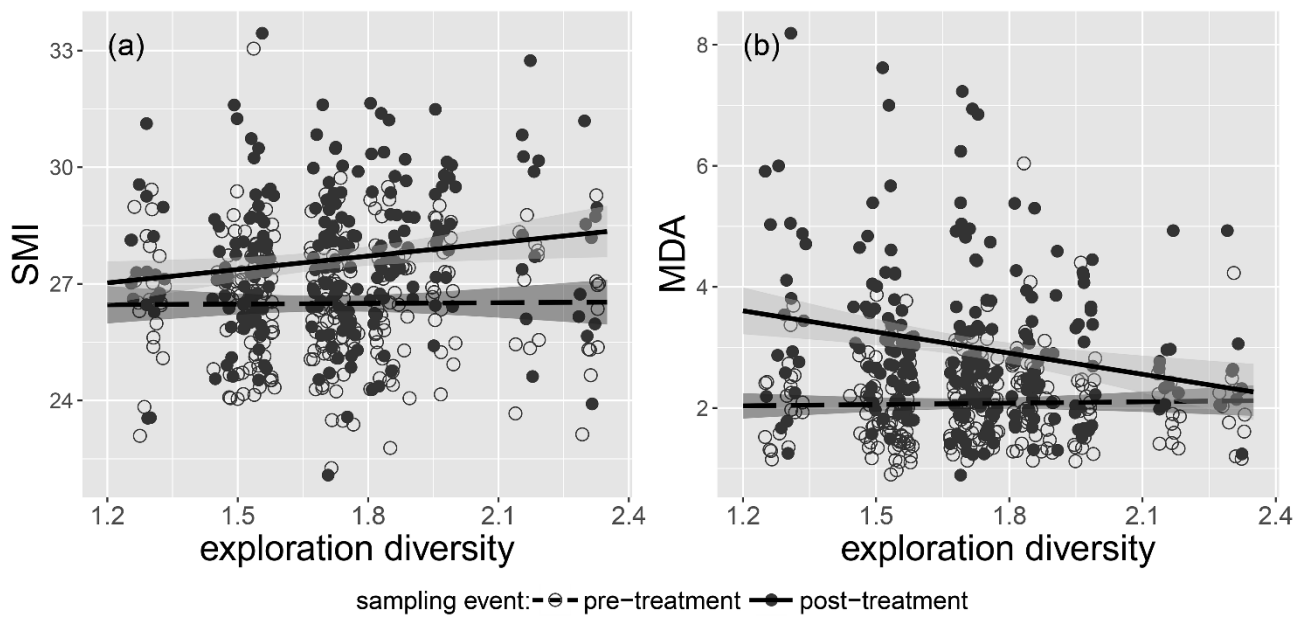
4) I do like the revised Figure 1 but encourage the authors to 1) make it colour-blind friendly and 2) put the raw (individual) data points on the figure as well. I also think the x-axis needs a label? I'd encourage them to put a line legend as an inset on Figure 2 as well. This would really help readers.

**4. Thank you for these suggestions. We modified both figures according to the Reviewer's suggestion, except adding the individual data points to figure 1.**

**We used the colour-blind friendly 'Dark2' colour palette from the R package 'RColorBrewer' on the new figure 1. We also added the "treatment" label to the x-axis. Although we agree that inclusion of the raw data in figure 1 might help some readers in understanding the data, we did not add the raw data points for the following reasons. First, figure 1 shows model predicted means  $\pm$  s.e. and, therefore, plotting the raw data are not trivial. Second, figure 1 has five plots (one plot for each physiological response variable) with eight bars (four treatment groups  $\times$  two sampling events), which makes the figure already complex. Adding the raw data points (240 dots per plot) would increase this complexity even more possibly to the detriment of readability or can possibly intimidate the reader to linger upon the figure. The modified figure 1:**



**We also added a point and line legend to figure 2 as suggested by the Reviewer. The modified figure 2:**



5) I thank the authors for the description regarding the papers authored by Pruitt that were cited here. Sounds like they vetted them and it is up to their discretion to cite them. It is unfortunate that a brief sentence couldn't describe this process given that other readers will not be aware this process occurred.

**5. We absolutely agree with the Reviewer and think one should be careful in citing papers authored by a scientist under investigation of misconduct. We've been careful and assessed whether the Pruitt papers we cite are suspected of containing fabricated data. Since yet there is no sign of cheating for the particular papers we cite, we objectively kept these papers in the reference list.**

**Thank you again for the constructive criticism. We are happy to see high-quality and objectively formulated reviews.**

Referee: 2

I read the revised manuscript on how flocks that vary in the composition of their group members' personality scores differ in their members' physiological parameters. The authors have addressed all my previously made (minor) suggestions and replied to my comments in a sufficient matter, by providing additional explanations in both the manuscript, the ESM, and in the authors' response letter.

**We are glad that the Reviewer appreciated our revision and gave further constructive comments.**

I only have very minor comments: One suggestion that may help improve the interpretability of figure 2: I would suggest introducing some scatter to the plot, because the open symbols are partly overlaid by the closed symbols. Also, maybe pairing open symbols with dashed lines and closed symbols with solid lines might be more intuitive than the current pairing.

**Figure 2: Thank you for these suggestions. We modified figure 2 according to the Reviewer's suggestions by adding a small amount of jitter to the data points to minimize their overlap, and we changed the pairing of symbols with the regression lines. Additionally, we added a point and line legend as requested by Reviewer #1. See the modified figure 2 inserted above.**

I saw the authors provided the raw data to their manuscript, which is well-prepared with annotations and explanations, and I would encourage the authors to also provide the script used for the data analysis, to further add to the transparency of the data analysis.

**Script used for data analyses: The Reviewer is right. We are also propagators of transparent science. Therefore, we deposited in Dryad all the data and analysis code (R script) supporting our results as suggested by the Reviewer. See lines #379–380.**

**Thank you again for the constructive criticism. We are happy to see high-quality and objectively formulated reviews. Below we inserted the revised manuscript with tracked changes.**