THE ROYAL SOCIETY PUBLISHING

PROCEEDINGS B

Psychoactive pollution suppresses individual differences in fish behaviour

Giovanni Polverino, Jake M. Martin, Michael G. Bertram, Vrishin R. Soman, Hung Tan, Jack A. Brand, Rachel T. Mason and Bob B. M. Wong

Article citation details
Proc. R. Soc. B **288**: 20202294.
http://dx.doi.org/10.1098/rspb.2020.2294

Review timeline

Original submission:	16 September 2020
Revised submission:	28 December 2020
Final acceptance:	15 January 2021

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

Review History

RSPB-2020-2294.R0 (Original submission)

Review form: Reviewer 1 (Raphaël Royauté)

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?
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.
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
Review of RSPB-2020-2294: "Psychoactive pollution suppresses individual differences in fish behaviour"

This manuscript focuses on how an important pharmaceutical contaminant affects fish behavior using a long-term exposure study. The authors show that differences among individuals decline in contaminant-exposed groups and that this effect seem to intensify at higher concentrations. The authors suggest that this pattern could be driven either by adaptation to contaminated environment resulting in loss of genetic diversity in exposed populations or because exposure to the contaminant early in life permanently alters the expression of behavioral variation. Overall, I quite enjoyed reading this article and find little to comment on regarding the study design and the behavioral trials conducted. I do have some issues with how the authors tested for differences in behavioral variance among treatments, however. First, I think the linear regression approach can be potentially misleading given that treatments concentrations are not spaced equally. This is important because non-linear dose responses abound in ecotoxicology and the study design is not well-equipped to tease apart linear from non-linear relationships with so few doses. There are also some contradictions between the model comparison provided in Table S1 (no effect of treatment on among-individual variance) and the inference based on the regression approach that would need to be addressed and investigated further. For these reasons, I think a simpler approach providing pairwise difference between treatments by variance components and repeatability would be much more appropriate. With these potential issues, I cannot completely trust the inference provided yet and make recommendations below that I hope the authors will find helpful.

Best,
Raphaël Royauté

Major comments:
1) The authors tested for differences in repeatability and among/within-individual differences in behavior by regressing posterior estimates against contamination treatments. This is problematic to me because it ignores that control and low-exposure groups are closer in their concentrations than the high-exposure group. The authors took care to take the distribution of errors into account by looping the procedure over the whole posterior distribution, but I worry that there are not enough dose classes here to properly test for linear vs. non-linear effects. I would much rather see the differences in repeatability/variance components tested in a pairwise manner. This can easily be done by subtracting all posterior repeatability estimates from the treated groups from

those of the control group: Delta R = R_Low – R_Control, R_High – R_Control and R_High – R_Low. You can compare whether the 95 % CIs overlap with 0 or calculate how many posterior estimates are of the same sign as the median, if you prefer to express the probability that the difference departs from 0. I think this would provide a much more appropriate test for the main hypothesis. See Royauté et al. 2015, White & Briffa 2017 or Tüzün et al. 2017 for examples in behavioral ecotoxicology, see also R code provided at the end.

2) The results presented in Table S1 seem in contradiction with the main inference. The model comparison indicates that there is essentially no difference between a model that fits heterogeneous variances by treatment vs. a model that ignores this difference. To me this is even more reason to test pairwise differences in variance and repeatability. There are several ways this can be done statistically, but, as explained above, I don't think the regression analysis used as the main inference tool is appropriate here.

3) Figure 2 also seems to contradict the argument that these changes in among-individual variance are linear, at least for activity. The among-individual variance is almost unchanged between low and high-exposure groups. It seems that any amount of fluoxetine present in the water is likely to result in these dramatic changes in behavioral expression. Provided these results hold up to additional analyses, this could be an interesting point for the discussion. Even minimal exposure, if sustained for long enough, results in fishes losing their personality differences.

4) The authors found shifts in both average and among-individual variance in risk-taking. Given that mean and variances are often related, it may also be appropriate to compare mean-standardized variances among treatments (coefficient of variation, or other appropriate metrics; see Hansen et al. 2011, Dochtermann & Royauté 2019 and Niemelä et al. 2019 for a practical example). This is not a major concern however!

5) Is there any evidence for individuals changing in plasticity with trial? I think adding a random slope model for trial in the model comparison in Table S1 would be useful. The way the model is coded, Trial 1 is set as the intercept so all random effects are estimated as deviation from this level. This is fine if there is little Trial x Individual plasticity, but would be important to check nonetheless.

6) I found the main result figure a little crowded and difficult to read. It would be good to show both the shifts in average behavior along with each level of behavioral variation. Color coding distributions and putting among and within-individual variance in separate panels would also help with readability. See also the tidybayes package for some great tools to plot results from MCMC models (http://mjskay.github.io/tidybayes/)

7) There are now quite a few studies that tested for differences in repeatability with exposure to contaminants. Is there a general consensus on which level of variation is most likely to be affected by exposure? This could be a useful addition to the discussion (space-permitting, of course!).

Line-by-line comments

LL231-232: "1,500,000 resamplings, 500,000 burn-ins, and 100 thinnings." Is that a typo for thinning? It usually doesn't make sense to try to get > 1,000 posterior samples from an MCMC run, unless you are combining over multiple chains. With a 100 thinnings, you would get 10,000 samples, which sounds overkill for this type of procedure. How do the models behave when using a longer thinning interval?

LL271-272: "We observed a linear decline…" I don't think you can really make this statement given that you are treating contamination levels as categorical (I think! This aspect was a bit unclear). In any case, the distance between the "high-exposure" and "low-exposure" groups is almost ten times greater that of the distance between the "low-exposure" group and the control. It doesn't really seem right to test for potential linear or non-linear effects the contaminant can

have on behavioral expression given that the study is not really designed for that. Which is completely fine in my opinion! I think it would be fair to speculate on these non-linearities in the discussion given that U-shape/inverted U-shape dose-response curves are common in ecotoxicology (see Hellou 2011 for example; but rarely investigated at the individual level, to my knowledge). I think it would be much better to focus on pairwise differences in repeatabilities/variance component to investigate these effects in greater details. This is a much more reliable approach in my opinion because metrics such as delta R (Rtreated – Rcontrol) are analogous to effect sizes.

LL283-284: "but we did not detect significant pairwise differences". Could you provide a Figure with the raw data or predicted values from the mixed models? Otherwise this paragraph is a little difficult to follow.

LL323-324: "Differences in behaviour between individuals are known to increase over their lifetime [50]". I don't think there is a strong consensus on whether state-dependent feedbacks are primarily positive or negative. Recent meta-analysis suggest little associations between state and behavior as well (Niemelä & Dingemanse 2018). My sense is that a negative feedback between risk taking and contaminant exposure is equally likely here. See also Montiglio & Royauté 2014 for some conceptual examples on how contaminant exposure can affect behavioral expression through multiple routes. This is just the article that is closest to mind, feel free to use any relevant review on the topic.

Figure 1: "Because of the nature of the risk-taking variable, high values represent low risk-taking." That doesn't really make sense because these y-axes represents repeability and variances. So higher values can only mean the trait is more repeatable (for the left-side panels). I think an additional figure showing the raw data by treatment would be very useful!

##
MCMCglmm code

```
#Prior for model 3 (all other priors would be variations around this formulation
prior = list(R=list(V=diag(2), nu=1.002), G=list(G1=list(V=diag(2), nu=1.002)))

# Model with equal among & within individual variances
mcmc.dist.m0 <- MCMCglmm(Dist.Moved.cm. ~ Fixed Effects, random=~ ID, other arguments)

# Model with heterogeneous among individual variance by treatment
mcmc.dist.m1 <- MCMCglmm(Dist.Moved.cm. ~ Fixed Effects, random=~idh(Treatment):ID,
other arguments)

# Model with heterogeneous residual variance (i.e. within-individual variance) by treatment
mcmc.dist.m2 <- MCMCglmm(Dist.Moved.cm. ~ Fixed Effects, random=~ ID, other arguments)

# Model with heterogeneous among-individual & residual variance
mcmc.dist.m3 <- MCMCglmm(Dist.Moved.cm. ~ Fixed Effects, random=~ idh(Treatment):ID,
rcov= idh(Treatment):units, other arguments)

# Model comparison
mcmc.dist.m0$DIC; mcmc.dist.m1$DIC; mcmc.dist.m2$DIC; mcmc.dist.m3$DIC;
# Pairwise differences can be computed by taking differences from the posterior distribution of #
random effect stored in mcmc.dist.m3$VCV

# See https://ecoevorxiv.org/tn7u5/ for more details and coding options
```

##

References

Dochtermann, N. A., & Royauté, R. (2019). The mean matters: going beyond repeatability to interpret behavioural variation. Animal Behaviour, 153, 147-150.

Hansen, T. F., Pélabon, C., & Houle, D. (2011). Heritability is not evolvability. Evolutionary Biology, 38(3), 258.

Hellou, J. (2011). Behavioural ecotoxicology, an "early warning" signal to assess environmental quality. Environmental Science and Pollution Research, 18(1), 1-11.

Montiglio, P. O., & Royauté, R. (2014). Contaminants as a neglected source of behavioural variation. Anim Behav, 88, 29-35.

Niemelä, P. T., & Dingemanse, N. J. (2018). Meta-analysis reveals weak associations between intrinsic state and personality. Proceedings of the Royal Society B: Biological Sciences, 285(1873), 20172823.

Niemelä, P. T., Niehoff, P. P., Gasparini, C., Dingemanse, N. J., & Tuni, C. (2019). Crickets become behaviourally more stable when raised under higher temperatures. Behavioral Ecology and Sociobiology, 73(6), 81.

Royauté, R., Buddle, C. M., & Vincent, C. (2015). Under the influence: sublethal exposure to an insecticide affects personality expression in a jumping spider. Functional Ecology, 29(7), 962-970.

Tüzün, N., Mueller, S., Koch, K., & Stoks, R. (2017). Pesticide-induced changes in personality depend on the urbanization level. Animal Behaviour, 134, 45-55.

White, S. J., & Briffa, M. (2017). How do anthropogenic contaminants (ACs) affect behaviour? Multi-level analysis of the effects of copper on boldness in hermit crabs. Oecologia, 183(2), 391-400.

Review form: Reviewer 2 (Mathilde Tissier)

Recommendation
Accept with minor revision (please list in comments)

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

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

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

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
The studied how long-term exposure to environmentally realistic and controlled concentrations of fluoxetine, a psychoactive pollutant, modifies the between and within individuals differences in behaviours. They used a sound and very well-designed experimental approach using mesocosms, and to provide very interesting and robust results. I really enjoyed reading this article, which I believe is very well written. In addition, I congratulate the authors for Figure 1 which speaks for itself. I mostly have minor comments and questions (though numerous) that you will find below, with the majority of the questions related to the methods. However, even if some of these questions are on the statistical analyses, I recognize that they can be answered quite easily and I don't believe that they would question the validity of the findings. My main comment would be on the discussion. The authors mostly discuss environmental applications of their results and especially the potential underlying mechanisms explaining the long-term effects of an exposure to fluoxetine on the variation in individual behaviour. This section, although interesting, could be shortened, as it goes beyond the scope of this study. This would leave more space to discuss the results. I would for instance appreciate to read more about the potential evolutionary consequences of a decrease in overall behavioural repeatability in activity at the population level in the wild, considering that there were no changes in individual plasticity. To me, this is at the core of your study/results, but the discussion is very brief on this. I provide more information below. I hope you will find my comments constructive.
With best regards,
Mathilde Tissier

Abstract
I would appreciate to see the sample size (L24) and a bit more on the (behavioural) methodology (tests conducted in an open-field and variable measured) as it is one of the central point of this study.

Introduction
L63-64 – wildlife? (instead of animals; since humans basically are animals).
L68 – I would suggest to include an additional reference here, which is one of the leading one in the field of inter-individual differences in animal behaviour and their consequences for wild populations and evolution Réale, D., Reader, S. M., Sol, D., McDougall, P. T., & Dingemanse, N. J. (2007). Integrating animal temperament within ecology and evolution. Biological reviews, 82(2), 291-318.
L87 – consider adding a brief statement on what these two papers brought to the study of pharmaceuticals or psychoactive contaminants effects on animal behaviour before highlighting

what they did not manage to achieve?
L116 – perhaps specify Figure 1c scenario 2 to guide the reader, which could have the reflex to look at scenario 1 (as it is mentioned first).

L118 – perhaps specify scenario 1?
> Suggestion: Why not displaying scenarios/panels on Figure 1 in the order mentioned in the text? It will smooth the reading.
> The third scenario is not mentioned. Can it be predicted? Perhaps you could mention that as scenarios 1 and 2 can both be predicted according to the literature and are non-mutually exclusive, there is a third prediction?

L122-126: this should be kept for your discussion.
Methods

L130-134 - I am sorry but I can't see the link with Figure 1a here? It does not show anything related to the sample size, sex ratio, mesocosms size or volume capacity? Perhaps consider replacing by a picture of the mesocosms, in the supplementary information if you're limited in the number of figures you can include inside this manuscript.

L149 – Supplementary Methods instead of Material?

L153-155 – something is bugging me here: you say that you sex only adults (thus after having determined the life stage) but then that your determination of the life-stage is based on the sex determination. Could you please specify? Maybe just rephrasing by stating that all individuals that measured more than 15mm were considered as adults and males were determined on the basis of their colouration and the presence of a fully developed gonopodium?

L157 – you mean 2L of water from its native mesocosm? Not native population (meaning in the wild)? I think so when looking at L142 and 161. But I would avoid the use of populations in your study and rather chose a more general terminology, such as group, treatment group or perhaps mesocosm population as in L161, and use it throughout the manuscript… to prevent any misunderstanding.

L165-166 – I am not sure that I follow you here. I think this part of the sentence is not necessary "and not exploring open spaces that are unfamiliar and potentially dangerous" as it gives the impression that your risk-taking variable is a combination of two variables (time spent in the refuge + exploration of open spaces yes/no). I think what you mean is that, when they are in the refuge, they consequently not explore risky unfamiliar spaces that may potentially dangerous? You can perhaps rephrase your sentence.

L190-192 – That's quite a nice sample size of individuals and repeatability for an experimental study that you have here.

L199 – why solely a visual inspection of the residuals of you models? Why didn't you conduct a test of normality on the residuals?

L204-208 – you did not test for any interaction between your fixed effects? To verify that all mesocosms from the same treatment show the same response for instance?

L211-213 – perhaps a bit more details would be helpful here? You should specify that you used the best likelihood ratio; and did you fix a ΔAIC under which you considered that your full and null models were equivalent? Why did you use both LRTs and ΔAIC to compare your models (you don't mention that you compare them, just that you use LRTs and ΔAIC between the full and reduced models)?

L221 – models (a s is missing)

L240 – what do you mean by classes? Age class? You say age above in your methods.

L250-251 – The Bonferroni is indeed very conservative, and has been criticized in ecology (I agree that each correction has been). Why not using a less conservative – though valid – approach such as the false discovery rate instead?

Results

L257-259 – I would keep that for the discussion rather than the result section.

L263 – 273 – 275 - Mentions to Figure 2: it would be helpful to know to which panel you are referring-to in the text.

Discussion

L293-299 – The two sentences are somehow redundant, although they also partially provide complementary information. I think they could be synthetized into one key finding. I wondered whether the sentence L293-297 could be deleted to focus only on the sentence L297-299 which highlights the key finding here. But perhaps we'd lose part of the information that you want to share here in doing so.

L315-317 – I agree with the statement but I believe that it should be supported by references. And perhaps consider spending a bit more of your discussion on this aspect. I think it is under-represented compared to the rest of your discussion, although it is directly linked to your results. Why are differences in behaviour important? In which specific situations this lack of differences could be detrimental for the population? Are you talking about individual fitness benefits? In that case, how does the fitness of an individual can be affected by the reduction of differences in behaviour in its population? And regarding the population, how can this reduction can have an effect? I think these are all fair points to discuss, and perhaps more related to your findings than what you discuss later-on. The proximate mechanisms behind is secondary to your study, as it goes beyond what your protocole allowed you to test.

L318-345 – Although I appreciate these two sections and find them interesting, they are beyond the scope of your study and could thus be reduced to provide more space to discuss your findings.

Figures
Fig 1: this figure is really appreciated! Perhaps just consider re-organizing the panels following the order of mention in the introduction (panel/scenario 2 is mentioned before scenario 1).

Fig 2: it is a bit complex to make the parallel between your figure 1 and your figure 2. I understand that you may be limited in representing the individual slopes and variations in figure 2 in the same way that you did in figure 1 given your sample size. But perhaps you could add the expected overall repeatability on the three panels of figure 1C to help the reader?

Also, perhaps consider adding letters as you did in figure 1 (a-d) for each panel and cite them in your result section.

Why not including the green to orange gradient shown on figure 1 below the X-axis?

These are just suggestions, as per se your figure is good. Maybe it is just that I put my expectations very high when looking at figure 1, and I would appreciate some visual parallel with your results. But this is me being picky, I admit.

Decision letter (RSPB-2020-2294.R0)

06-Nov-2020

Dear Dr Polverino:

Your manuscript has now been peer reviewed and the reviews have been assessed by an Associate Editor. The reviewers' comments (not including confidential comments to the Editor) and the comments from the Associate Editor are included at the end of this email for your reference. As you will see, the reviewers and the Editors have raised some concerns with your manuscript and we would like to invite you to revise your manuscript to address them.

We do not allow multiple rounds of revision so we urge you to make every effort to fully address all of the comments at this stage. If deemed necessary by the Associate Editor, your manuscript will be sent back to one or more of the original reviewers for assessment. If the original reviewers are not available we may invite new reviewers. Please note that we cannot guarantee eventual acceptance of your manuscript at this stage.

To submit your revision please 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 Revision". Your manuscript number has been appended to denote a revision.

When submitting your revision please upload a file under "Response to Referees" - in the "File Upload" section. This should document, point by point, how you have responded to the reviewers' and Editors' comments, and the adjustments you have made to the 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.

Your main manuscript should be submitted as a text file (doc, txt, rtf or tex), not a PDF. Your figures should be submitted as separate files and not included within the main manuscript file.

When revising your manuscript you should also ensure that it adheres to our editorial policies (https://royalsociety.org/journals/ethics-policies/). You should pay particular attention to the following:

Research ethics:
If your study contains research on humans please ensure that you detail in the methods section whether you obtained ethical approval from your local research ethics committee and gained informed consent to participate from each of the participants.

Use of animals and field studies:
If your study uses animals please include details in the methods section of any approval and licences given to carry out the study and include full details of how animal welfare standards were ensured. Field studies should be conducted in accordance with local legislation; please include details of the appropriate permission and licences that you obtained to carry out the field work.

Data accessibility and data citation:
It is a condition of publication that you make available the data and research materials supporting the results in the article. Please see our Data Sharing Policies (https://royalsociety.org/journals/authors/author-guidelines/#data). Datasets should be deposited in an appropriate publicly available repository and details of the associated accession number, link or DOI to the datasets must be included in the Data Accessibility section of the

article (https://royalsociety.org/journals/ethics-policies/data-sharing-mining/). Reference(s) to datasets should also be included in the reference list of the article with DOIs (where available).

In order to ensure effective and robust dissemination and appropriate credit to authors the dataset(s) used should also be fully cited and listed in the references.

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), 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.

For more information please see our open data policy http://royalsocietypublishing.org/data-sharing.

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. Please try to submit all supplementary material as a single file.

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].

Please submit a copy of your revised paper within three weeks. If we do not hear from you within this time your manuscript will be rejected. If you are unable to meet this deadline please let us know as soon as possible, as we may be able to grant a short extension.

Thank you for submitting your manuscript to Proceedings B; we look forward to receiving your revision. If you have any questions at all, please do not hesitate to get in touch.

Best wishes,
Dr Daniel Costa
mailto: proceedingsb@royalsociety.org

Associate Editor
Board Member: 1
Comments to Author:
Two expert reviewers have evaluated the MS and were enthusiastic about the overall study design and presentation. Both reviewers made excellent specific observations throughout the MS regarding elements that can be improved and further refined. Of particular importance, one reviewer notes some modeling choices that could limit or even mislead the interpretation of the data and provides detailed suggestions for alternative statistical approaches that can be checked for robustness of the conclusions.

Reviewer(s)' Comments to Author:
Referee: 1
Comments to the Author(s)
Review of RSPB-2020-2294: “Psychoactive pollution suppresses individual differences in fish behaviour”

This manuscript focuses on how an important pharmaceutical contaminant affects fish behavior using a long-term exposure study. The authors show that differences among individuals decline in contaminant-exposed groups and that this effect seem to intensify at higher concentrations. The authors suggest that this pattern could be driven either by adaptation to contaminated environment resulting in loss of genetic diversity in exposed populations or because exposure to the contaminant early in life permanently alters the expression of behavioral variation. Overall, I quite enjoyed reading this article and find little to comment on regarding the study design and the behavioral trials conducted. I do have some issues with how the authors tested for differences in behavioral variance among treatments, however. First, I think the linear regression approach can be potentially misleading given that treatments concentrations are not spaced equally. This is important because non-linear dose responses abound in ecotoxicology and the study design is not well-equipped to tease apart linear from non-linear relationships with so few doses. There are also some contradictions between the model comparison provided in Table S1 (no effect of treatment on among-individual variance) and the inference based on the regression approach that would need to be addressed and investigated further. For these reasons, I think a simpler approach providing pairwise difference between treatments by variance components and repeatability would be much more appropriate. With these potential issues, I cannot completely trust the inference provided yet and make recommendations below that I hope the authors will find helpful.

Best,
Raphaël Royauté

Major comments:
1) The authors tested for differences in repeatability and among/within-individual differences in behavior by regressing posterior estimates against contamination treatments. This is problematic to me because it ignores that control and low-exposure groups are closer in their concentrations than the high-exposure group. The authors took care to take the distribution of errors into account by looping the procedure over the whole posterior distribution, but I worry that there are not enough dose classes here to properly test for linear vs. non-linear effects. I would much rather see the differences in repeatability/variance components tested in a pairwise manner. This can easily be done by subtracting all posterior repeatability estimates from the treated groups from those of the control group: Delta R = R_Low – R_Control, R_High – R_Control and R_High – R_Low. You can compare whether the 95 % CIs overlap with 0 or calculate how many posterior estimates are of the same sign as the median, if you prefer to express the probability that the difference departs from 0. I think this would provide a much more appropriate test for the main hypothesis. See Royauté et al. 2015, White & Briffa 2017 or Tüzün et al. 2017 for examples in behavioral ecotoxicology, see also R code provided at the end.

2) The results presented in Table S1 seem in contradiction with the main inference. The model comparison indicates that there is essentially no difference between a model that fits heterogeneous variances by treatment vs. a model that ignores this difference. To me this is even more reason to test pairwise differences in variance and repeatability. There are several ways this can be done statistically, but, as explained above, I don't think the regression analysis used as the main inference tool is appropriate here.

3) Figure 2 also seems to contradict the argument that these changes in among-individual variance are linear, at least for activity. The among-individual variance is almost unchanged between low and high-exposure groups. It seems that any amount of fluoxetine present in the water is likely to result in these dramatic changes in behavioral expression. Provided these results hold up to additional analyses, this could be an interesting point for the discussion. Even minimal exposure, if sustained for long enough, results in fishes losing their personality differences.

4) The authors found shifts in both average and among-individual variance in risk-taking. Given that mean and variances are often related, it may also be appropriate to compare mean-

standardized variances among treatments (coefficient of variation, or other appropriate metrics; see Hansen et al. 2011, Dochtermann & Royauté 2019 and Niemelä et al. 2019 for a practical example). This is not a major concern however!

5) Is there any evidence for individuals changing in plasticity with trial? I think adding a random slope model for trial in the model comparison in Table S1 would be useful. The way the model is coded, Trial 1 is set as the intercept so all random effects are estimated as deviation from this level. This is fine if there is little Trial x Individual plasticity, but would be important to check nonetheless.

6) I found the main result figure a little crowded and difficult to read. It would be good to show both the shifts in average behavior along with each level of behavioral variation. Color coding distributions and putting among and within-individual variance in separate panels would also help with readability. See also the tidybayes package for some great tools to plot results from MCMC models (http://mjskay.github.io/tidybayes/)

7) There are now quite a few studies that tested for differences in repeatability with exposure to contaminants. Is there a general consensus on which level of variation is most likely to be affected by exposure? This could be a useful addition to the discussion (space-permitting, of course!).

Line-by-line comments

LL231-232: "1,500,000 resamplings, 500,000 burn-ins, and 100 thinnings." Is that a typo for thinning? It usually doesn't make sense to try to get > 1,000 posterior samples from an MCMC run, unless you are combining over multiple chains. With a 100 thinnings, you would get 10,000 samples, which sounds overkill for this type of procedure. How do the models behave when using a longer thinning interval?

LL271-272: "We observed a linear decline…" I don't think you can really make this statement given that you are treating contamination levels as categorical (I think! This aspect was a bit unclear). In any case, the distance between the "high-exposure" and "low-exposure" groups is almost ten times greater that of the distance between the "low-exposure" group and the control. It doesn't really seem right to test for potential linear or non-linear effects the contaminant can have on behavioral expression given that the study is not really designed for that. Which is completely fine in my opinion! I think it would be fair to speculate on these non-linearities in the discussion given that U-shape/inverted U-shape dose-response curves are common in ecotoxicology (see Hellou 2011 for example; but rarely investigated at the individual level, to my knowledge). I think it would be much better to focus on pairwise differences in repeatabilities/variance component to investigate these effects in greater details. This is a much more reliable approach in my opinion because metrics such as delta R (Rtreated – Rcontrol) are analogous to effect sizes.

LL283-284: "but we did not detect significant pairwise differences". Could you provide a Figure with the raw data or predicted values from the mixed models? Otherwise this paragraph is a little difficult to follow.

LL323-324: "Differences in behaviour between individuals are known to increase over their lifetime [50]". I don't think there is a strong consensus on whether state-dependent feedbacks are primarily positive or negative. Recent meta-analysis suggest little associations between state and behavior as well (Niemelä & Dingemanse 2018). My sense is that a negative feedback between risk taking and contaminant exposure is equally likely here. See also Montiglio & Royauté 2014 for some conceptual examples on how contaminant exposure can affect behavioral expression through multiple routes. This is just the article that is closest to mind, feel free to use any relevant review on the topic.

Figure 1: "Because of the nature of the risk-taking variable, high values represent low risk-taking." That doesn't really make sense because these y-axes represents repeability and variances. So higher values can only mean the trait is more repeatable (for the left-side panels). I think an additional figure showing the raw data by treatment would be very useful!

```
#########################################################################
MCMCglmm code

#Prior for model 3 (all other priors would be variations around this formulation
prior = list(R=list(V=diag(2), nu=1.002), G=list(G1=list(V=diag(2), nu=1.002)))

# Model with equal among & within individual variances
mcmc.dist.m0 <- MCMCglmm(Dist.Moved.cm. ~ Fixed Effects, random=~ ID, other arguments)

# Model with heterogeneous among individual variance by treatment
mcmc.dist.m1 <- MCMCglmm(Dist.Moved.cm. ~ Fixed Effects, random=~idh(Treatment):ID,
other arguments)

# Model with heterogeneous residual variance (i.e. within-individual variance) by treatment
mcmc.dist.m2 <- MCMCglmm(Dist.Moved.cm. ~ Fixed Effects, random=~ ID, other arguments)

# Model with heterogeneous among-individual & residual variance
mcmc.dist.m3 <- MCMCglmm(Dist.Moved.cm. ~ Fixed Effects, random=~ idh(Treatment):ID,
rcov= idh(Treatment):units, other arguments)

# Model comparison
mcmc.dist.m0$DIC; mcmc.dist.m1$DIC; mcmc.dist.m2$DIC; mcmc.dist.m3$DIC;

# Pairwise differences can be computed by taking differences from the posterior distribution of #
random effect stored in mcmc.dist.m3$VCV
# See https://ecoevorxiv.org/tn7u5/ for more details and coding options
#########################################################################
```

References

Dochtermann, N. A., & Royauté, R. (2019). The mean matters: going beyond repeatability to interpret behavioural variation. Animal Behaviour, 153, 147-150.

Hansen, T. F., Pélabon, C., & Houle, D. (2011). Heritability is not evolvability. Evolutionary Biology, 38(3), 258.

Hellou, J. (2011). Behavioural ecotoxicology, an "early warning" signal to assess environmental quality. Environmental Science and Pollution Research, 18(1), 1-11.

Montiglio, P. O., & Royauté, R. (2014). Contaminants as a neglected source of behavioural variation. Anim Behav, 88, 29-35.

Niemelä, P. T., & Dingemanse, N. J. (2018). Meta-analysis reveals weak associations between intrinsic state and personality. Proceedings of the Royal Society B: Biological Sciences, 285(1873), 20172823.

Niemelä, P. T., Niehoff, P. P., Gasparini, C., Dingemanse, N. J., & Tuni, C. (2019). Crickets become behaviourally more stable when raised under higher temperatures. Behavioral Ecology and Sociobiology, 73(6), 81.

Royauté, R., Buddle, C. M., & Vincent, C. (2015). Under the influence: sublethal exposure to an insecticide affects personality expression in a jumping spider. Functional Ecology, 29(7), 962-970.

Tüzün, N., Mueller, S., Koch, K., & Stoks, R. (2017). Pesticide-induced changes in personality depend on the urbanization level. Animal Behaviour, 134, 45-55.

White, S. J., & Briffa, M. (2017). How do anthropogenic contaminants (ACs) affect behaviour? Multi-level analysis of the effects of copper on boldness in hermit crabs. Oecologia, 183(2), 391-400.

Referee: 2
Comments to the Author(s)
The studied how long-term exposure to environmentally realistic and controlled concentrations of fluoxetine, a psychoactive pollutant, modifies the between and within individuals differences in behaviours. They used a sound and very well-designed experimental approach using mesocosms, and to provide very interesting and robust results. I really enjoyed reading this article, which I believe is very well written. In addition, I congratulate the authors for Figure 1 which speaks for itself. I mostly have minor comments and questions (though numerous) that you will find below, with the majority of the questions related to the methods. However, even if some of these questions are on the statistical analyses, I recognize that they can be answered quite easily and I don't believe that they would question the validity of the findings. My main comment would be on the discussion. The authors mostly discuss environmental applications of their results and especially the potential underlying mechanisms explaining the long-term effects of an exposure to fluoxetine on the variation in individual behaviour. This section, although interesting, could be shortened, as it goes beyond the scope of this study. This would leave more space to discuss the results. I would for instance appreciate to read more about the potential evolutionary consequences of a decrease in overall behavioural repeatability in activity at the population level in the wild, considering that there were no changes in individual plasticity. To me, this is at the core of your study/results, but the discussion is very brief on this. I provide more information below. I hope you will find my comments constructive.
With best regards,
Mathilde Tissier

Abstract
I would appreciate to see the sample size (L24) and a bit more on the (behavioural) methodology (tests conducted in an open-field and variable measured) as it is one of the central point of this study.

Introduction
L63-64 – wildlife? (instead of animals; since humans basically are animals).
L68 – I would suggest to include an additional reference here, which is one of the leading one in the field of inter-individual differences in animal behaviour and their consequences for wild populations and evolution Réale, D., Reader, S. M., Sol, D., McDougall, P. T., & Dingemanse, N. J. (2007). Integrating animal temperament within ecology and evolution. Biological reviews, 82(2), 291-318.
L87 – consider adding a brief statement on what these two papers brought to the study of pharmaceuticals or psychoactive contaminants effects on animal behaviour before highlighting what they did not manage to achieve?
L116 – perhaps specify Figure 1c scenario 2 to guide the reader, which could have the reflex to look at scenario 1 (as it is mentioned first).

L118 – perhaps specify scenario 1?
> Suggestion: Why not displaying scenarios/panels on Figure 1 in the order mentioned in the text? It will smooth the reading.

> The third scenario is not mentioned. Can it be predicted? Perhaps you could mention that as scenarios 1 and 2 can both be predicted according to the literature and are non-mutually exclusive, there is a third prediction?

L122-126: this should be kept for your discussion.
Methods

L130-134 - I am sorry but I can't see the link with Figure 1a here? It does not show anything related to the sample size, sex ratio, mesocosms size or volume capacity? Perhaps consider replacing by a picture of the mesocosms, in the supplementary information if you're limited in the number of figures you can include inside this manuscript.

L149 – Supplementary Methods instead of Material?

L153-155 – something is bugging me here: you say that you sex only adults (thus after having determined the life stage) but then that your determination of the life-stage is based on the sex determination. Could you please specify? Maybe just rephrasing by stating that all individuals that measured more than 15mm were considered as adults and males were determined on the basis of their colouration and the presence of a fully developed gonopodium?

L157 – you mean 2L of water from its native mesocosm? Not native population (meaning in the wild)? I think so when looking at L142 and 161. But I would avoid the use of populations in your study and rather chose a more general terminology, such as group, treatment group or perhaps mesocosm population as in L161, and use it throughout the manuscript… to prevent any misunderstanding.

L165-166 – I am not sure that I follow you here. I think this part of the sentence is not necessary "and not exploring open spaces that are unfamiliar and potentially dangerous" as it gives the impression that your risk-taking variable is a combination of two variables (time spent in the refuge + exploration of open spaces yes/no). I think what you mean is that, when they are in the refuge, they consequently not explore risky unfamiliar spaces that may potentially dangerous? You can perhaps rephrase your sentence.

L190-192 – That's quite a nice sample size of individuals and repeatability for an experimental study that you have here.

L199 – why solely a visual inspection of the residuals of you models? Why didn't you conduct a test of normality on the residuals?

L204-208 – you did not test for any interaction between your fixed effects? To verify that all mesocosms from the same treatment show the same response for instance?

L211-213 – perhaps a bit more details would be helpful here? You should specify that you used the best likelihood ratio; and did you fix a ΔAIC under which you considered that your full and null models were equivalent? Why did you use both LRTs and ΔAIC to compare your models (you don't mention that you compare them, just that you use LRTs and ΔAIC between the full and reduced models)?

L221 – models (a s is missing)

L240 – what do you mean by classes? Age class? You say age above in your methods.

L250-251 – The Bonferroni is indeed very conservative, and has been criticized in ecology (I agree that each correction has been). Why not using a less conservative – though valid – approach such as the false discovery rate instead?

Results

L257-259 – I would keep that for the discussion rather than the result section.

L263 – 273 – 275 - Mentions to Figure 2: it would be helpful to know to which panel you are referring-to in the text.

Discussion

L293-299 – The two sentences are somehow redundant, although they also partially provide complementary information. I think they could be synthetized into one key finding. I wondered whether the sentence L293-297 could be deleted to focus only on the sentence L297-299 which highlights the key finding here. But perhaps we'd lose part of the information that you want to share here in doing so.

L315-317 – I agree with the statement but I believe that it should be supported by references. And perhaps consider spending a bit more of your discussion on this aspect. I think it is under-represented compared to the rest of your discussion, although it is directly linked to your results. Why are differences in behaviour important? In which specific situations this lack of differences could be detrimental for the population? Are you talking about individual fitness benefits? In that case, how does the fitness of an individual can be affected by the reduction of differences in behaviour in its population? And regarding the population, how can this reduction can have an effect? I think these are all fair points to discuss, and perhaps more related to your findings than what you discuss later-on. The proximate mechanisms behind is secondary to your study, as it goes beyond what your protocole allowed you to test.

L318-345 – Although I appreciate these two sections and find them interesting, they are beyond the scope of your study and could thus be reduced to provide more space to discuss your findings.

Figures
Fig 1: this figure is really appreciated! Perhaps just consider re-organizing the panels following the order of mention in the introduction (panel/scenario 2 is mentioned before scenario 1).

Fig 2: it is a bit complex to make the parallel between your figure 1 and your figure 2. I understand that you may be limited in representing the individual slopes and variations in figure 2 in the same way that you did in figure 1 given your sample size. But perhaps you could add the expected overall repeatability on the three panels of figure 1C to help the reader?
Also, perhaps consider adding letters as you did in figure 1 (a-d) for each panel and cite them in your result section.
Why not including the green to orange gradient shown on figure 1 below the X-axis?
These are just suggestions, as per se your figure is good. Maybe it is just that I put my expectations very high when looking at figure 1, and I would appreciate some visual parallel with your results. But this is me being picky, I admit.

Author's Response to Decision Letter for (RSPB-2020-2294.R0)

See Appendix A.

Decision letter (RSPB-2020-2294.R1)

15-Jan-2021

Dear Dr Polverino

I am pleased to inform you that your manuscript entitled "Psychoactive pollution suppresses individual differences in fish behaviour" 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

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.

Your article has been estimated as being 10 pages long. Our Production Office will be able to confirm the exact length at proof stage.

Paper charges
An e-mail request for payment of any related charges will be sent out after proof stage (within approximately 2-6 weeks). 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.

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,
Dr Daniel Costa
Editor, Proceedings B
mailto: proceedingsb@royalsociety.org

Associate Editor:
Comments to Author:
Thank you for your thorough and thoughtful responses to the reviews. Overall, the work is well-conducted and touches upon a number of interesting areas of biological significance. With the post-review improvements, I believe that this paper will be of value to a general scientific audience.

Giovanni Polverino, PhD
Centre for Evolutionary Biology
School of Animal Biology (M092)
Perth, WA, 6009, Australia

28 December 2020

Appendix A

Dear Prof. Costa,

We are pleased to submit a revised version of our manuscript RSPB-2020-2294 'Psychoactive pollution suppresses individual differences in fish behaviour' for consideration in *Proceedings B*.

Below is a point-by-point response detailing how we have addressed the Associate Editor and reviewers' comments. Their input has greatly strengthened the manuscript, and we thank them for their constructive insights and suggestions.

We look forward to a decision in due course.

Sincerely yours,

Giovanni Polverino, Jake M. Martin, Michael G. Bertram, Vrishin R. Soman, Hung Tan, Jack A. Brand, Rachel T. Mason, and Bob B.M. Wong

Associate Editor

COMMENT
Two expert reviewers have evaluated the MS and were enthusiastic about the overall study design and presentation. Both reviewers made excellent specific observations throughout the MS regarding elements that can be improved and further refined. Of particular importance, one reviewer notes some modeling choices that could limit or even mislead the interpretation of the data and provides detailed suggestions for alternative statistical approaches that can be checked for robustness of the conclusions.

RESPONSE
We are delighted to have received such positive and enthusiastic feedback, and thank the Editor and the two reviewers for their input. We have addressed each of the reviewers' comments in the point-by-point responses below. In particular, following reviewer 1's excellent suggestions, we have included a second statistical approach in our revised analysis. Importantly, the new analysis supports the original findings, and reinforces the robustness of our key conclusions.

THE UNIVERSITY OF WESTERN AUSTRALIA

Reviewer #1 (Raphaël Royauté)

COMMENT #1

This manuscript focuses on how an important pharmaceutical contaminant affects fish behavior using a long-term exposure study. The authors show that differences among individuals decline in contaminant-exposed groups and that this effect seem to intensify at higher concentrations. The authors suggest that this pattern could be driven either by adaptation to contaminated environment resulting in loss of genetic diversity in exposed populations or because exposure to the contaminant early in life permanently alters the expression of behavioral variation. Overall, I quite enjoyed reading this article and find little to comment on regarding the study design and the behavioral trials conducted. I do have some issues with how the authors tested for differences in behavioral variance among treatments, however. First, I think the linear regression approach can be potentially misleading given that treatments concentrations are not spaced equally. This is important because non-linear dose responses abound in ecotoxicology and the study design is not well-equipped to tease apart linear from non-linear relationships with so few doses. There are also some contradictions between the model comparison provided in Table S1 (no effect of treatment on among-individual variance) and the inference based on the regression approach that would need to be addressed and investigated further. For these reasons, I think a simpler approach providing pairwise difference between treatments by variance components and repeatability would be much more appropriate. With these potential issues, I cannot completely trust the inference provided yet and make recommendations below that I hope the authors will find helpful.

RESPONSE

We thank the reviewer for his positive response and appreciate his insightful and detailed comments, which we have addressed point-by-point in the responses below. We are particularly grateful for the reviewer's statistical suggestions, which have now been incorporated into the manuscript (see details in responses to comments #2-4, 6-7, and 11 from reviewer 1).

COMMENT #2

The authors tested for differences in repeatability and among/within-individual differences in behavior by regressing posterior estimates against contamination treatments. This is problematic to me because it ignores that control and low-exposure groups are closer in their concentrations than the high-exposure group. The authors took care to take the distribution of errors into account by looping the procedure over the whole posterior distribution, but I worry that there are not enough dose classes here to properly test for linear vs. non-linear effects. I would much rather see the differences in repeatability/variance components tested in a pairwise manner. This can easily be done by subtracting all posterior repeatability estimates from the treated groups from those of the control group: Delta R = R_Low – R_Control, R_High – R_Control and R_High – R_Low. You can compare whether the 95 % CIs overlap with 0 or calculate how many posterior estimates are of the same sign as the median, if you prefer to express the probability that the difference departs from 0. I think this would provide a much more appropriate test for the main hypothesis. See Royauté et al. 2015, White & Briffa 2017 or Tüzün et al. 2017 for examples in behavioral ecotoxicology, see also R code provided at the end.

RESPONSE

Done. As suggested, we have now carried out pairwise comparisons in our revised analysis, and tested whether variance components (between and within individuals) and repeatability estimates differed between treatments. The results of these analyses support our original finding — fluoxetine reduces the repeatability of individual behaviour by eroding variation between but not within individuals — and suggest that very low doses of the pollutant are enough to observe dramatic changes in the behavioural expression of fish. We believe that these new statistics are complementary to our initial approach, revealing further details about the effects of fluoxetine on

the behavioural variation observed. We note that our previous analysis did already account for treatment levels which are not equally spaced (see lines 236-241 and Table S2 in the revised text, in which we have tested both continuous and categorical predictors), and has been often used in the personality literature to test for assumed linear relationships with three levels of treatment variation (see for example [1, 2]). So our revised text builds on the evidence from both statistical approaches, offering a solid and more comprehensive picture of the effects of a psychoactive pollutant on individual variation in animal behaviour. The revised sections of the MS now read, as follows:

Lines 237-245: ".. a linear regression was fitted to both variance components and repeatability scores that had been randomly sampled from the posterior distributions using treatment (unexposed, low fluoxetine, high fluoxetine) as a continuous predictor. We also tested treatment as a categorical predictor to check for, and confirm, the consistency of results from the linear regressions (see Table S2). We repeated this procedure 10,000 times. Statistical significance was inferred from the empirical distribution of the 10,000 slopes. In addition, we performed pairwise comparisons between variance components and repeatability estimates across treatments, as suggested by [3], and assumed significant differences when the 95% confidence intervals did not overlap with zero."

Lines 278-284: "We observed a decline in behavioural differences between fish in activity levels at increasing concentrations of pharmaceutical pollution, with dramatic reductions in the behavioural expression of fish already observed at the lowest concentration of the pollutant (Figure 2 and Table S3). By contrast, the variation in activity levels did not differ within individuals across treatments. Neither variation within nor between individuals differed across treatments for risk-taking (Figure 2 and Table S3)."

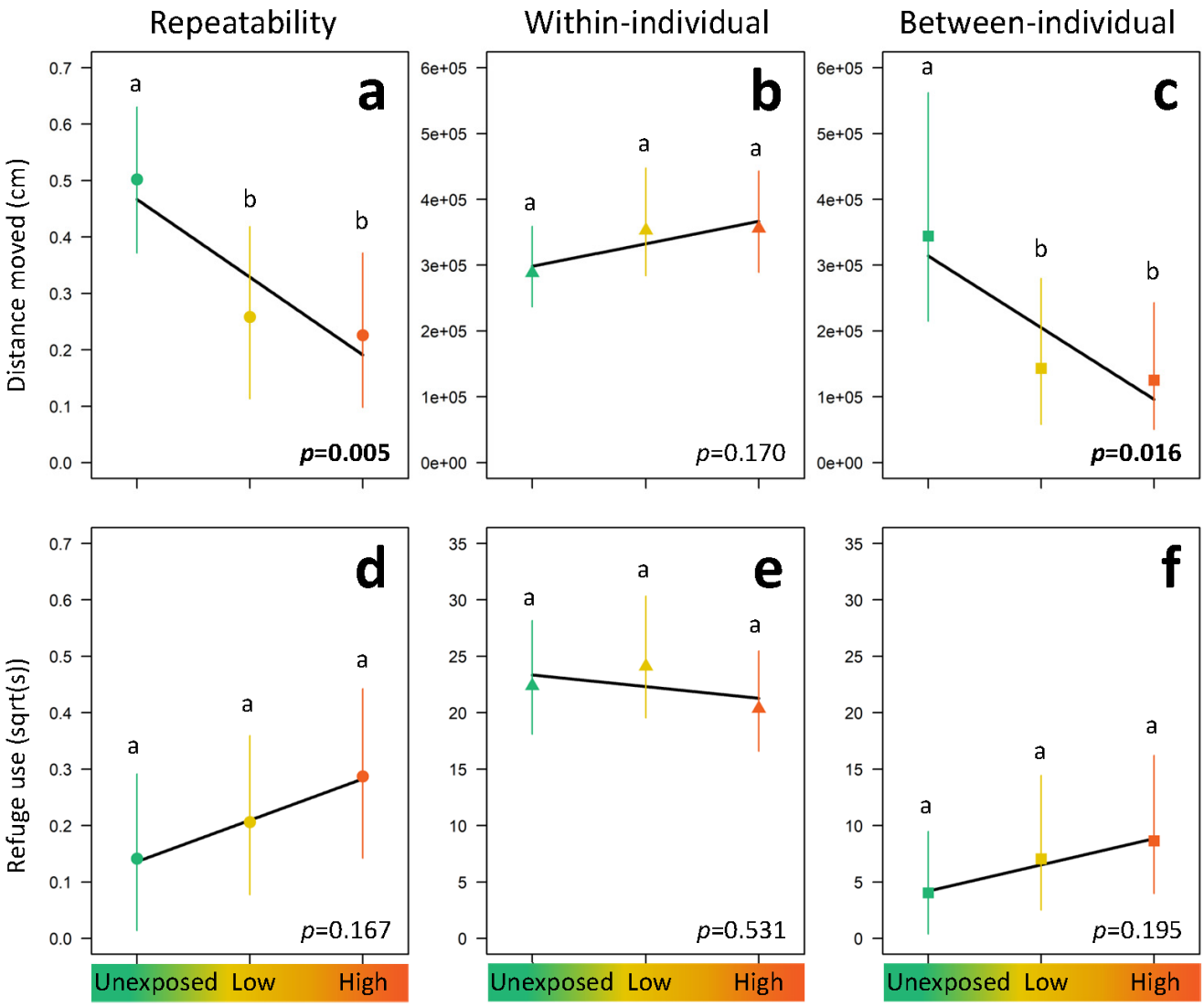
Lines 305-309: ".. behavioural repeatability declines with increasing concentrations of the pollutant, and a substantial drop in repeatability already occurs at the lowest dose. Specifically, long-term exposure to fluoxetine pollution erodes behavioural variation between but not within individuals, uncovering the evolutionary process behind the overall decline."

The University of Western Australia
M000, Perth WA 6009 Australia

T +61 8 6488 1473

E giovanni.polverino@uwa.edu.au
CRICOS Provider Code 00126G

Figure 2: ***Change in the overall measure of repeatability and its variance components (within and between individuals) for activity (distance moved) and risk-taking (refuge use) across treatments (unexposed, low fluoxetine, high fluoxetine).*** *Changes in the overall measure of repeatability (**a** and **d**; medians as circles), and in behavioural variation within (**b** and **e**; medians as triangles) and between (**c** and **f**; medians as squares) individuals. Vertical lines represent 95% credible intervals, and estimates not sharing a common superscript are significantly different.*



The University of Western Australia
M000, Perth WA 6009 Australia
T +61 8 6488 1473
E giovanni.polverino@uwa.edu.au
CRICOS Provider Code 00126G

Table S3: ***Pairwise comparisons between variance components (within and between individuals) and repeatability estimates for activity (distance moved) and risk-taking (refuge use) across treatments (unexposed, low fluoxetine, high fluoxetine).*** *Mean estimates and their 95% credible intervals are represented, and significant results (bold) correspond to estimated differences whose credible intervals do not overlap with zero.*

Models	Pairwise comparisons	Estimate	95% CIs
Activity: distance moved			
$V_{between}$	Unexposed-Low	**172168.100**	**12196.090, 429603.500**
	Unexposed-High	**233493.500**	**38073.480, 432233.400**
	Low-High	–537.279	–130613.300, 164493
V_{within}	Unexposed-Low	–61931.780	–166754.400, 39138.950
	Unexposed-High	–59393.510	–167617.200, 32214.990
	Low-High	10665.240	–109928.300, 112270
Repeatability	Unexposed-Low	**0.256**	**0.038, 0.442**
	Unexposed-High	**0.279**	**0.068, 0.447**
	Low-High	0.009	–0.170, 0.232
Risk-taking: refuge use			
$V_{between}$	Unexposed-Low	–2.681	–10.650, 4.539
	Unexposed-High	–5.280	–12.814, 2.638
	Low-High	–0.181	–10.425, 6.890
V_{within}	Unexposed-Low	–1.756	–9.229, 5.535
	Unexposed-High	2.278	–4.803, 8.573
	Low-High	3.554	–3.109, 10.870
Repeatability	Unexposed-Low	–0.059	–0.266, 0.132
	Unexposed-High	–0.188	–0.339, 0.064
	Low-High	–0.070	–0.285, 0.124

COMMENT #3

The results presented in Table S1 seem in contradiction with the main inference. The model comparison indicates that there is essentially no difference between a model that fits heterogeneous variances by treatment vs. a model that ignores this difference. To me this is even more reason to test pairwise differences in variance and repeatability. There are several ways this can be done statistically, but, as explained above, I don't think the regression analysis used as the main inference tool is appropriate here.

RESPONSE

Great pick up! The contradiction mentioned by the reviewer was the result of a typo. We have now included all fixed effects — which were previously missing — into the models summarised in Table S1, and the results support our main finding. We have amended the MS, as follows:

The University of Western Australia
M000, Perth WA 6009 Australia
T +61 8 6488 1473
E giovanni.polverino@uwa.edu.au
CRICOS Provider Code 00126G

Lines 271-274: ".. allowing individual slopes to vary between treatments improved the model fit for activity—confirming heterogeneous variance across treatments— but not risk-taking, while individual adjustments were comparable over time and across age and sex in both behavioural traits (Table S1)."

Table S1: ***Comparison of LMMs with different random (co)variance structures for activity (distance moved) and risk-taking (refuge use)****. We compared two models with both random intercepts (ID) and slopes (treatment, class, or trial) across individuals against a reduced model with only random intercepts (ID). We also compared the model with random intercepts (ID) with a model in which random intercepts were excluded. Significant improvements in the model fit are tested with both Akaike information criteria (ΔAIC) and likelihood-ratio tests (p). Treatment (unexposed, low fluoxetine, high fluoxetine), class (juveniles, males, females), mesocosm population (four per treatment), trial (four repeated measures per individual), and time of day are included as fixed effects in all models. Significance was set at α < 0.05 and significant results are in bold; the model with random intercepts and treatment as the random slope was best supported as the most parsimonious model for activity, while only random intercepts but not random slopes explained a significant portion of the variance in risk-taking. The random structure is written in the syntax of the lme4 R package.*

Model	χ^2_1	ΔAIC	*p*
Activity: distance moved			
(1|ID)	118.040	116.04	**< 0.001**
(treatment|ID)	5.998	2.000	**0.049**
(class|ID)	4.814	< 0.001	0.090
(trial|ID)	1.424	2.600	0.491
Risk-taking: refuge use			
(1|ID)	45.081	43.081	**< 0.001**
(treatment|ID)	1.080	2.900	0.583
(class|ID)	0.323	3.700	0.851
(trial|ID)	3.353	0.700	0.187

COMMENT #4

Figure 2 also seems to contradict the argument that these changes in among-individual variance are linear, at least for activity. The among-individual variance is almost unchanged between low and high-exposure groups. It seems that any amount of fluoxetine present in the water is likely to result in these dramatic changes in behavioral expression. Provided these results hold up to additional analyses, this could be an interesting point for the discussion. Even minimal exposure, if sustained for long enough, results in fishes losing their personality differences.

RESPONSE

The reviewer is correct. As mentioned above, we have now included pairwise comparisons between variance components (between and within individuals) and repeatability estimates across treatments, and confirmed that the pollutant has dramatic effects already at very small doses. Accordingly, we have revised our statistics, figures, tables, and the main text to include this

new evidence (see details above for how we have revised both statistics and results in response to comment #2 from the reviewer). We have also revised our Discussion to focus on this interesting aspect pointed out by the reviewer. The amended text reads, as follows:

Lines 310-313: "The key finding is that chronic exposure to even very low concentrations of fluoxetine erodes variation in activity levels between individuals, but not plasticity within individuals; reduced variation between individuals drives decline in overall behavioural variation (repeatability)."

Lines 333-336: "By collapsing the diversity in the behavioural strategies among individuals, chronic exposure to even very small concentrations of this psychoactive drug has the potential of reducing resilience [4] and compromise the adaptive capacity [5] of animal populations to survive in a rapidly changing world [6-8]."

COMMENT #5
The authors found shifts in both average and among-individual variance in risk-taking. Given that mean and variances are often related, it may also be appropriate to compare mean-standardized variances among treatments (coefficient of variation, or other appropriate metrics; see Hansen et al. 2011, Dochtermann & Royauté 2019 and Niemelä et al. 2019 for a practical example). This is not a major concern however!

RESPONSE
We agree that this might be an additional and interesting aspect to consider. However, as this is not a major concern for the reviewer, and because the information is tangential to our main findings (i.e. between-individual variance does not vary across treatments for risk-taking), we have opted not to include the additional analyses and text. Importantly, this also means that we can continue to focus the narrative on the main findings without being constrained by the strict word limit for Proc B.

COMMENT #6
Is there any evidence for individuals changing in plasticity with trial? I think adding a random slope model for trial in the model comparison in Table S1 would be useful. The way the model is coded, Trial 1 is set as the intercept so all random effects are estimated as deviation from this level. This is fine if there is little Trial x Individual plasticity, but would be important to check nonetheless.

RESPONSE
Done. The new analysis supports our assumption that random slopes for trial (i.e. plasticity) do not explain a significant portion of the variance, both for activity and risk talking. The revised text now reads, as follows:

Lines 215-224: "In preliminary analyses, we also tested for the presence of heterogeneous variance between treatments, classes (age and sex), and trials (that is, random slopes/regression). We did this by running two separate models for each behavioural trait in which we included random intercepts and allowed individual slopes to either vary between treatments, classes, or trials. There was no evidence that inclusion of class and trial as random slopes increased model fit—no differences were detected between models with random intercepts and slopes or only intercepts via LRTs and ΔAIC—and we did not retain these terms in our final models (Table S1). Instead, comparisons between models identified heterogeneous individual variance across treatments in activity, but not in risk-taking (Table S1)."

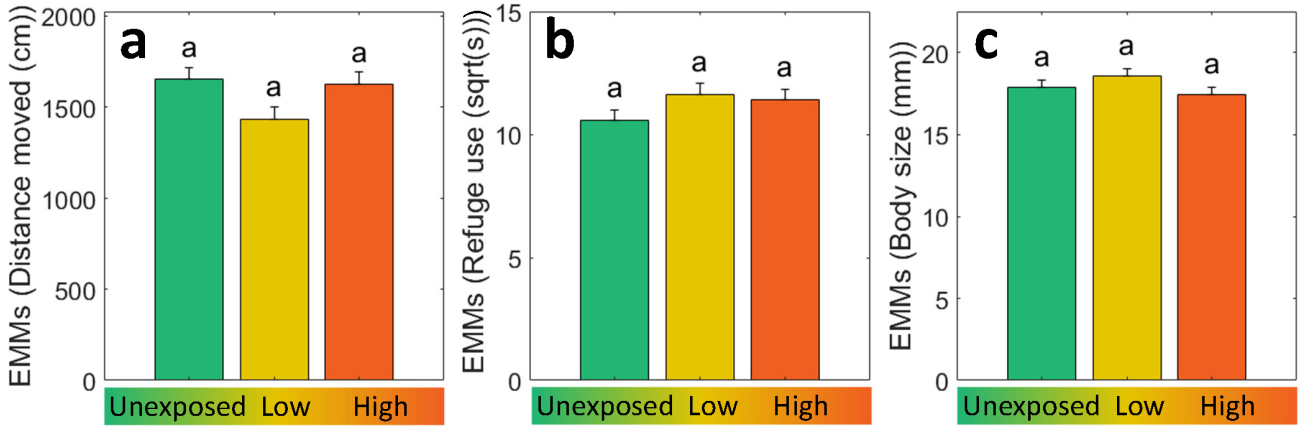
Table s1: (presented above in the response to comment #3 from the reviewer)

COMMENT #7
I found the main result figure a little crowded and difficult to read. It would be good to show both the shifts in average behavior along with each level of behavioral variation. Color coding distributions and putting among and within-individual variance in separate panels would also help with readability. See also the tidybayes package for some great tools to plot results from MCMC models (http://mjskay.github.io/tidybayes/)

RESPONSE
Done. Reviewer 2 raised similar points to improve the clarity of the figure. We have now revised Figure 2 according to the suggestions of both reviewers. In particular, we have used different colours for the three treatments, we have included a legend on the x axes — as per Figure 1 — and placed between- and within-individual variance for activity and risk taking into separate panels. We have also included superscripts to distinguish estimates which are significantly different (repeatability, and behavioural variation within and between individuals). Results on shifts in average traits have been presented in the Supplementary Material (Figure S2), and also include mean variation in body size. The relevant changes are, as follows:

Figure 2: (presented above in the response to comment #2 from the reviewer)

Figure S2: ***Pairwise differences in activity (a; distance moved), risk-taking (b; refuge use), and (c) body size of fish across treatments (unexposed, low fluoxetine, high fluoxetine).*** *Estimated marginal means (EMMs + SE) account for the contribution of the other fixed and random effects included in each model: class (juveniles, males, females), mesocosm population (four per treatment), trial (four repeated measures per individual), time of day, and random intercepts (individual ID) for activity and risk-taking; and class and random intercepts (mesocosm population ID) for body size. We ran pairwise comparisons with the conservative Bonferroni method. Significance was set at $\alpha < 0.05$, and means not sharing a common superscript are significantly different. Because of the nature of the risk-taking variable, high values represent low risk-taking.*



COMMENT #8
There are now quite a few studies that tested for differences in repeatability with exposure to contaminants. Is there a general consensus on which level of variation is most likely to be affected by exposure? This could be a useful addition to the discussion (space-permitting, of course!).

The University of Western Australia
M000, Perth WA 6009 Australia
T +61 8 6488 1473
E giovanni.polverino@uwa.edu.au
CRICOS Provider Code 00126G

RESPONSE
As the reviewer correctly highlights, there is a small but growing literature on the effects of contaminants on behavioural variation (from heavy metals to radioactive contaminants; for two relevant reviews on this topic, see [9, 10]). But, to the best of our knowledge, there are very few studies available in the context of pharmaceutical pollutants to be able to make any generalizable conclusions at this stage without being overly speculative. Indeed, it's important to point out that even fewer studies have attempted to disentangle the role of behavioural variation between and within individuals, which was the primary motivation for the current study. For these reasons and to save space, we prefer not to add an extra section on these issues in our Discussion.

COMMENT #9
LL231-232: "1,500,000 resamplings, 500,000 burn-ins, and 100 thinnings." Is that a typo for thinning? It usually doesn't make sense to try to get > 1,000 posterior samples from an MCMC run, unless you are combining over multiple chains. With a 100 thinnings, you would get 10,000 samples, which sounds overkill for this type of procedure. How do the models behave when using a longer thinning interval?

RESPONSE
We have re-run our models with a longer interval (10 thinning) and verified that the new models still mixed and converged properly, yielding to repeatability and variance estimates identical to our previous ones to the third decimal point. Since results are unchanged, we have decided to maintain a 100 thinning interval, as suggested by Dingemanse and Dochtermann for quantifying individual variation in behaviour with a mixed-effect approach [11].

COMMENT #10
LL271-272: "We observed a linear decline…" I don't think you can really make this statement given that you are treating contamination levels as categorical (I think! This aspect was a bit unclear). In any case, the distance between the "high-exposure" and "low-exposure" groups is almost ten times greater that of the distance between the "low-exposure" group and the control. It doesn't really seem right to test for potential linear or non-linear effects the contaminant can have on behavioral expression given that the study is not really designed for that. Which is completely fine in my opinion! I think it would be fair to speculate on these non-linearities in the discussion given that U-shape/inverted U-shape dose-response curves are common in ecotoxicology (see Hellou 2011 for example; but rarely investigated at the individual level, to my knowledge). I think it would be much better to focus on pairwise differences in repeatabilities/variance component to investigate these effects in greater details. This is a much more reliable approach in my opinion because metrics such as delta R (Rtreated – Rcontrol) are analogous to effect sizes.

RESPONSE
As detailed earlier, in response to the reviewer's comments on this point, we have included a new statistical approach in our analysis, revising the presentation of our findings (Figure 2 and Table S3), and adopting a more conservative approach for describing the dose-response relationship observed (see details above in our responses to comments #2 and 4 from the reviewer). Regarding the specific comment raised by the reviewer here, we have revised the text, as follows:

Lines 278-282: "We observed a decline in behavioural differences between fish in activity levels at increasing concentrations of pharmaceutical pollution, with dramatic reductions in the behavioural expression of fish already observed at the lowest concentration of the pollutant (Figure 2 and Table S3)."

The University of Western Australia
M000, Perth WA 6009 Australia
T +61 8 6488 1473
E giovanni.polverino@uwa.edu.au
CRICOS Provider Code 00126G

COMMENT #11
LL283-284: "but we did not detect significant pairwise differences". Could you provide a Figure with the raw data or predicted values from the mixed models? Otherwise this paragraph is a little difficult to follow.

RESPONSE
Done. We have included a new supplementary figure in our revised manuscript (Figure S2), which provides the predicted values from each mixed model and the pairwise comparisons between treatments.

Figure S2: (presented above in the response to comment #7 from the reviewer)

COMMENT #12
LL323-324: "Differences in behaviour between individuals are known to increase over their lifetime [50]". I don't think there is a strong consensus on whether state-dependent feedbacks are primarily positive or negative. Recent meta-analysis suggest little associations between state and behavior as well (Niemelä & Dingemanse 2018). My sense is that a negative feedback between risk taking and contaminant exposure is equally likely here. See also Montiglio & Royauté 2014 for some conceptual examples on how contaminant exposure can affect behavioral expression through multiple routes. This is just the article that is closest to mind, feel free to use any relevant review on the topic.

RESPONSE
This is a fair point. We note that a negative (or positive) feedback between risk taking and contaminant exposure is unlikely here, since we did not find significant (individual-level) variation in risk-taking between unexposed and fluoxetine-exposed fish. Nevertheless, we have amended this entire paragraph in the revised manuscript to be more careful with our wording, and included the key references suggested by the reviewer. The revised section now reads, as follows:

Lines 342-349: "Recent evidence suggests that differences in behaviour between individuals can increase over their lifetime (see [1] and references therein), relying on positive feedback loops between behavioural tendencies early in life and the environment [12]. For example, individuals that are more active, explorative, and risk-prone early in life should be more competitive in securing resources and should grow faster [13] — but see [14] — with cumulative life experiences over time (e.g. winning or losing when competing for resources) resulting in larger differences in behavioural strategies in adulthood [15]."

Lines 365-370: "This evidence adds to a growing body of literature, in which exposure to environmental contaminants has been suggested to alter behavioural expression in animals through multiple routes [9], and even strengthen behavioural variation (e.g. when the probability to encounter contaminated environments varies between individuals with different activity levels, furtherly increasing their difference in the future)."

COMMENT #13
Figure 1: "Because of the nature of the risk-taking variable, high values represent low risk- taking." That doesn't really make sense because these y-axes represents repeability and variances. So higher values can only mean the trait is more repeatable (for the left-side panels). I think an additional figure showing the raw data by treatment would be very useful!

RESPONSE
Thanks for pointing this out. We have removed this text from the revised figure legend (Figure 2). Also, as detailed above in our responses to comments #7 and 11, we have included a new figure

(Figure S2) and revised the text of the manuscript to provide more details about our data-by-treatment.

Reviewer #2 (Mathilde Tissier)

COMMENT #1
The studied how long-term exposure to environmentally realistic and controlled concentrations of fluoxetine, a psychoactive pollutant, modifies the between and within individuals differences in behaviours. They used a sound and very well-designed experimental approach using mesocosms, and to provide very interesting and robust results. I really enjoyed reading this article, which I believe is very well written. In addition, I congratulate the authors for Figure 1 which speaks for itself. I mostly have minor comments and questions (though numerous) that you will find below, with the majority of the questions related to the methods. However, even if some of these questions are on the statistical analyses, I recognize that they can be answered quite easily and I don't believe that they would question the validity of the findings. My main comment would be on the discussion. The authors mostly discuss environmental applications of their results and especially the potential underlying mechanisms explaining the long-term effects of an exposure to fluoxetine on the variation in individual behaviour. This section, although interesting, could be shortened, as it goes beyond the scope of this study. This would leave more space to discuss the results. I would for instance appreciate to read more about the potential evolutionary consequences of a decrease in overall behavioural repeatability in activity at the population level in the wild, considering that there were no changes in individual plasticity. To me, this is at the core of your study/results, but the discussion is very brief on this. I provide more information below. I hope you will find my comments constructive.

RESPONSE
We thank the reviewer for her positive feedback and suggestions. In particular, the reviewer raises an excellent point about expanding our discussion regarding the potential evolutionary consequences of lowered personality variation in wild fish populations. We have now devoted an entire section to this point, as follows:

Lines 327-336: "In any case, unexpressed behavioural differences between individuals are likely to diminish their fitness benefits and, over time, reduce the magnitude of variation observed at the population level, as observed in fish populations under different selective regimes [16-18] and predicted for wildlife under anthropogenic disturbance [4]. For instance, risk-prone individuals would secure no more resources than risk-averse ones if their behavioural variation is consistently suppressed by fluoxetine, reducing the intensity of selection for maintaining such variation. By collapsing the diversity in the behavioural strategies among individuals, chronic exposure to even very small concentrations of this psychoactive drug has the potential of reducing resilience [4] and compromise the adaptive capacity [5] of animal populations to survive in a rapidly changing world [6-8]."

COMMENT #2
Abstract: I would appreciate to see the sample size (L24) and a bit more on the (behavioural) methodology (tests conducted in an open-field and variable measured) as it is one of the central point of this study.

RESPONSE
Done. As suggested, we have added the requested information to the Abstract, as follows:

Lines 25-27: "Fish (unexposed: n = 59, low fluoxetine: n = 57, high fluoxetine: n = 58) were repeatedly assayed on four separate occasions for activity and risk-taking behaviour."

We did not use the term "open-field assay" in the Abstract because it appears to have a very specific meaning in the field of behavioural ecology. Specifically, open-field assays are most often used to describe willingness to explore a circular open field with a uniform white bottom. This is different from the assay used in our experiment, where fish were tested in a square arena with a white bottom but with a squared dark region in one corner to serve as a refuge. Therefore, to avoid confusion in the description of the experimental protocol in the Abstract, we have avoided including this terminology and, instead, refer specifically to just the kind of behaviours assayed.

COMMENT #3
L63-64 – wildlife? (instead of animals; since humans basically are animals).

RESPONSE
Done. The revised text reads, as follows:

Lines 63-65: "So it remains largely unknown whether pollution by psychoactive drugs impacts some individuals more than others [10, 19], and attenuates individual differences in wildlife as it does in humans [20]."

COMMENT #4
L68 – I would suggest to include an additional reference here, which is one of the leading one in the field of inter-individual differences in animal behaviour and their consequences for wild populations and evolution Réale, D., Reader, S. M., Sol, D., McDougall, P. T., & Dingemanse, N. J. (2007). Integrating animal temperament within ecology and evolution. Biological reviews, 82(2), 291-318.

RESPONSE
Done. This key reference was already cited throughout the manuscript, and it's now also included in the sentence mentioned by the reviewer.

COMMENT #5
L87 – consider adding a brief statement on what these two papers brought to the study of pharmaceuticals or psychoactive contaminants effects on animal behaviour before highlighting what they did not manage to achieve?

RESPONSE
Done. The new section reads, as follows:

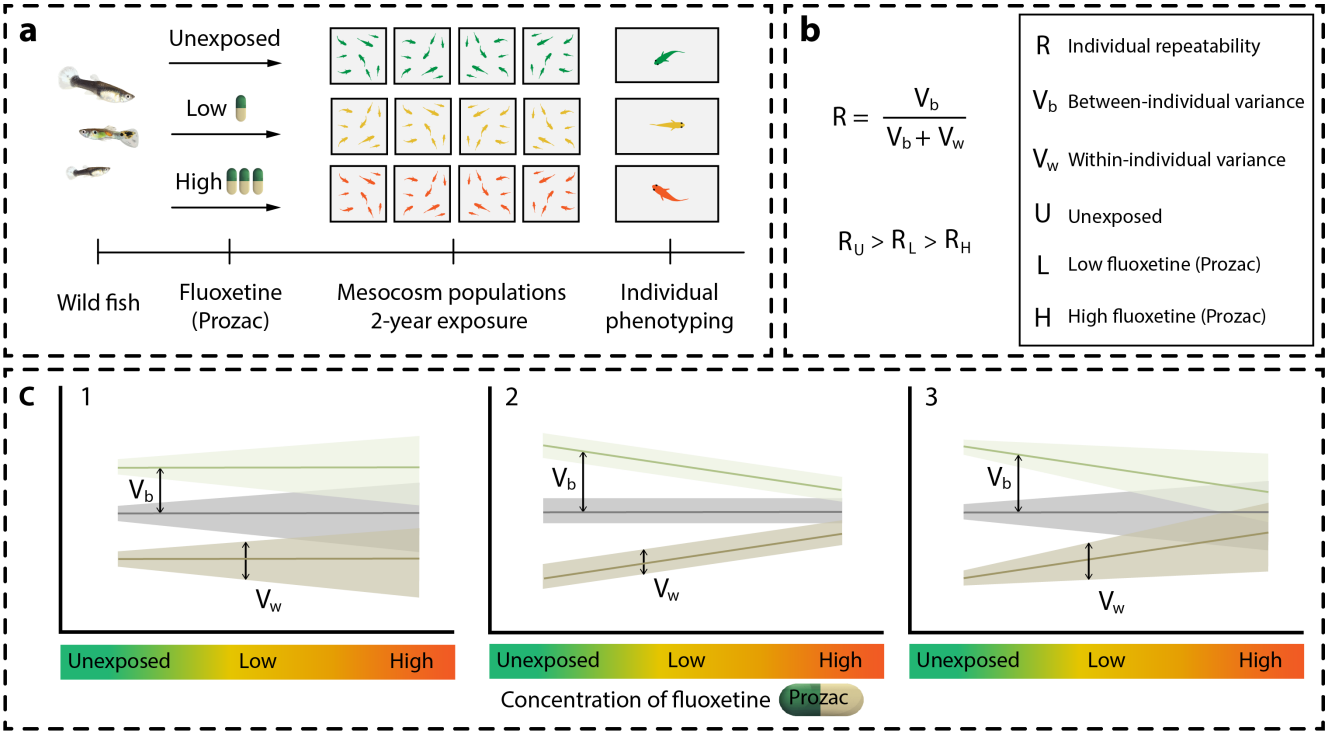
Lines 87-91: "While a few studies have shown that behavioural changes do occur at the individual level in animals exposed to contaminants [21, 22], none have mounted a comprehensive, systematic investigation of the ecological and evolutionary effects of psychoactive contaminants on individual variation in animal behaviour."

COMMENT #6
L116 – perhaps specify Figure 1c scenario 2 to guide the reader, which could have the reflex to look at scenario 1 (as it is mentioned first).

RESPONSE
Done. We have reversed the order of panels 1 and 2, so that panels are now described in the text in the same order in which they appear in the figure, as follows:

The University of Western Australia
M000, Perth WA 6009 Australia
T +61 8 6488 1473
E giovanni.polverino@uwa.edu.au
CRICOS Provider Code 00126G

Figure 1: ***Schematic of the experimental design.*** ***(a)*** *Wild fish (juveniles, and adult males and females) from a rainforest-fed stream in northern Australia—free from fluoxetine contamination—were split into 12 independent mesocosm populations. After five months fluoxetine exposure started, and lasted for 24 months. Coloured guppy symbols represent each treatment group (mean ± SE): green (unexposed), yellow (low fluoxetine, 40 ± 3 ng/L), and red (high fluoxetine, 366 ± 28 ng/L). For all treatments we assessed individual behavioural phenotypes.* ***(b)*** *Repeatability of activity levels and risk-taking was calculated as the proportion of within- and between-individual variance. We expected repeatability to be inversely associated with the level of fluoxetine exposure.* ***(c)*** *Behavioural responses of three individuals over an environmental gradient (fluoxetine concentration): behavioural variation within and between individuals is represented as the thickness of dark regions surrounding each line and the distance between lines, respectively. The predicted decline in repeatability with increasing fluoxetine concentration can be explained through three possible scenarios: (1) a decrease in between-individual differences, (2) an increase in within-individual differences, (2) a decrease in between-individual differences, and (3) a simultaneous decrease in between-individual differences and increase in within-individual differences.*



We have also adjusted the text, as follows:

Lines 118-123: "So we might expect a fish to adjust its individual behaviour when challenged by exposure to fluoxetine pollution (increasing within-individual variation; Figure 1c panel 1). Alternatively, we might expect fish to behave more similarly to each other because fluoxetine is formulated to narrow behavioural extremes [20] (decreasing between-individual variation; Figure 1c panel 2), or a combination of both scenarios (increasing within-individual variation and decreasing between-individual variation; Figure 1c panel 3)."

COMMENT #7
L118 – perhaps specify scenario 1?

> Suggestion: Why not displaying scenarios/panels on Figure 1 in the order mentioned in the text? It will smooth the reading.
> The third scenario is not mentioned. Can it be predicted? Perhaps you could mention that as scenarios 1 and 2 can both be predicted according to the literature and are non-mutually exclusive, there is a third prediction?

RESPONSE
Done. As detailed above, we have revised the figure to display the panels in the same order as described in the text. We have also included a description for scenario 3, as suggested by the reviewer, as follows:

Lines 122-123: ".. or a combination of both scenarios (increasing within-individual variation and decreasing between-individual variation; Figure 1c panel 3)."

COMMENT #8
L122-126: this should be kept for your discussion.

RESPONSE
Done. We have removed this passage from the Introduction of the revised manuscript.

COMMENT #9
L130-134 - I am sorry but I can't see the link with Figure 1a here? It does not show anything related to the sample size, sex ratio, mesocosms size or volume capacity? Perhaps consider replacing by a picture of the mesocosms, in the supplementary information if you're limited in the number of figures you can include inside this manuscript.

RESPONSE
Fair point. We have deleted the reference to Figure 1a from this sentence. Given space constraints, we opted not to include a picture of the actual mesocosms as it would not contribute anything of value to the manuscript beyond what has already been described in the text (i.e. 12 rectangular metal tanks).

COMMENT #10
L149 – Supplementary Methods instead of Material?

RESPONSE
Done.

COMMENT #11
L153-155 – something is bugging me here: you say that you sex only adults (thus after having determined the life stage) but then that your determination of the life-stage is based on the sex determination. Could you please specify? Maybe just rephrasing by stating that all individuals that measured more than 15mm were considered as adults and males were determined on the basis of their colouration and the presence of a fully developed gonopodium?

RESPONSE
Done. We have amended this section in the revised text to improve clarity, as follows:

Lines 154-156: "Each fish was visually inspected to determine its life stage. Females over 15 mm and males displaying colouration and a fully developed gonopodium were considered adults [23]."

The University of Western Australia
M000, Perth WA 6009 Australia
T +61 8 6488 1473
E giovanni.polverino@uwa.edu.au
CRICOS Provider Code 00126G

COMMENT #12
L157 – you mean 2L of water from its native mesocosm? Not native population (meaning in the wild)? I think so when looking at L142 and 161. But I would avoid the use of populations in your study and rather chose a more general terminology, such as group, treatment group or perhaps mesocosm population as in L161, and use it throughout the manuscript… to prevent any misunderstanding.

RESPONSE
Thanks for the suggestion. We have fixed this aspect to avoid confusion, and referred to mesocosm populations throughout the revised manuscript.

COMMENT #13
L165-166 – I am not sure that I follow you here. I think this part of the sentence is not necessary "and not exploring open spaces that are unfamiliar and potentially dangerous" as it gives the impression that your risk-taking variable is a combination of two variables (time spent in the refuge + exploration of open spaces yes/no). I think what you mean is that, when they are in the refuge, they consequently not explore risky unfamiliar spaces that may potentially dangerous? You can perhaps rephrase your sentence.

RESPONSE
Done. We have rephrased this section to improve clarity. The text now reads, as follows:

Lines 165-168: "We chose activity level (distance moved in cm) and risk-taking (use of the refuge in seconds, and consequently not exploring open spaces that are unfamiliar and potentially dangerous) as reference traits [24, 25]."

COMMENT #14
L190-192 – That's quite a nice sample size of individuals and repeatability for an experimental study that you have here.

RESPONSE
Thank you. As suggested by the reviewer in her comment #2, we have now also added this information into the Abstract.

COMMENT #15
L199 – why solely a visual inspection of the residuals of you models? Why didn't you conduct a test of normality on the residuals?

RESPONSE
Visually inspecting traces of the sampled output and density estimates is, to the best of our knowledge, the standard approach to examine the model residuals within a Bayesian framework — as the Markov chain Monte Carlo (MCMC) method used here. This approach is largely adopted in the literature (see for example [26]), and it is sufficient for assessing the correct distribution of the models residuals (see details in the "MCMCglmm Course Notes" Figure 1.11, https://cran.r-project.org/web/packages/MCMCglmm/vignettes/CourseNotes.pdf).

COMMENT #16
L204-208 – you did not test for any interaction between your fixed effects? To verify that all mesocosms from the same treatment show the same response for instance?

The University of Western Australia
M000, Perth WA 6009 Australia
T +61 8 6488 1473
E giovanni.polverino@uwa.edu.au
CRICOS Provider Code 00126G

RESPONSE
We did not have specific a priori predictions for testing interactions between fixed effects in these treatment-specific models, since we already had evidence that such interactions were not significant when tested in our general models (see Table 1). Also, we have already tested for potential differences between mesocosm populations from the same treatment group — which are not present — by including mesocosm population as a fixed effect in our treatment-specific models (see Table S5).

COMMENT #17
L211-213 – perhaps a bit more details would be helpful here? You should specify that you used the best likelihood ratio; and did you fix a ΔAIC under which you considered that your full and null models were equivalent? Why did you use both LRTs and ΔAIC to compare your models (you don't mention that you compare them, just that you use LRTs and ΔAIC between the full and reduced models)?

RESPONSE
Done. We have revised the text to clarify this aspect, which now reads, as follows:

Lines 212-215: "To do this, we used both likelihood ratio tests (LRTs) and Akaike information criteria (ΔAIC) to compare a full model, in which individual identities (random intercepts) were present, with a null model, in which random intercepts were excluded; we chose models with the best likelihood ratio, and lower AIC."

COMMENT #18
L221 – models (a s is missing)

RESPONSE
Thanks for pointing this out. We have fixed it.

COMMENT #19
L240 – what do you mean by classes? Age class? You say age above in your methods.

RESPONSE
We have revised the text to clarify this point, which now reads, as follows:

Lines 246-248: "We tested whether fish from different treatments (unexposed, low fluoxetine, high fluoxetine) and classes (juveniles, males, females) differed on average in their activity, risk-taking, and body size, by fitting a comprehensive LMM separately for each of these traits."

COMMENT #20
L250-251 – The Bonferroni is indeed very conservative, and has been criticized in ecology (I agree that each correction has been). Why not using a less conservative – though valid – approach such as the false discovery rate instead?

RESPONSE
We appreciate the reviewer's comment. The important point here is that our findings were robust, even when applying a more conservative (but still widely used) approach.

COMMENT #21
L257-259 – I would keep that for the discussion rather than the result section.

The University of Western Australia
M000, Perth WA 6009 Australia
T +61 8 6488 1473
E giovanni.polverino@uwa.edu.au
CRICOS Provider Code 00126G

RESPONSE
Done. We have removed this section from the revised manuscript.

COMMENT #22
L263 – 273 – 275 - Mentions to Figure 2: it would be helpful to know to which panel you are referring-to in the text.

RESPONSE
Done. We have numbered the panels in the revised figure, and referred to the specific panels throughout the text.

COMMENT #23
L293-299 – The two sentences are somehow redundant, although they also partially provide complementary information. I think they could be synthetized into one key finding. I wondered whether the sentence L293-297 could be deleted to focus only on the sentence L297-299 which highlights the key finding here. But perhaps we'd lose part of the information that you want to share here in doing so.

RESPONSE
The reviewer is correct in her conclusion about the potential loss of vital information if the two sentences were combined. While the first sentence refers to the overall reduction in behavioural repeatability caused by fish being exposed to fluoxetine, the second sentence explains the underlying mechanism which drives such decline, i.e. fluoxetine erodes variation between, but not within, individuals. These are two distinct results, and therefore we preferred to maintain the sentences as they are.

COMMENT #24
L315-317 – I agree with the statement but I believe that it should be supported by references. And perhaps consider spending a bit more of your discussion on this aspect. I think it is under-represented compared to the rest of your discussion, although it is directly linked to your results. Why are differences in behaviour important? In which specific situations this lack of differences could be detrimental for the population? Are you talking about individual fitness benefits? In that case, how does the fitness of an individual can be affected by the reduction of differences in behaviour in its population? And regarding the population, how can this reduction can have an effect? I think these are all fair points to discuss, and perhaps more related to your findings than what you discuss later-on. The proximate mechanisms behind is secondary to your study, as it goes beyond what your protocole allowed you to test.

RESPONSE
Thank you for the excellent suggestion. As detailed above, we have now expanded on this aspect and revised the discussion by including a new paragraph, which reads, as follows:

Lines 327-336: "In any case, unexpressed behavioural differences between individuals are likely to diminish their fitness benefits and, over time, reduce the magnitude of variation observed at the population level, as observed in fish populations under different selective regimes [16-18] and predicted for wildlife under anthropogenic disturbance [4]. For instance, risk-prone individuals would secure no more resources than risk-averse ones if their behavioural variation is consistently suppressed by fluoxetine, reducing the intensity of selection for maintaining such variation. By collapsing the diversity in the behavioural strategies among individuals, chronic exposure to even very small concentrations of this psychoactive drug has the potential of reducing resilience [4] and compromise the adaptive capacity [5] of animal populations to survive in a rapidly changing world [6-8]."

The University of Western Australia
M000, Perth WA 6009 Australia
T +61 8 6488 1473
E giovanni.polverino@uwa.edu.au
CRICOS Provider Code 00126G

We also note that the ecological importance of intra-population variation in behaviour is discussed extensively earlier in the text (see lines 66-85 of the revised manuscript), so we preferred not repeating that information in full here to save space.

COMMENT #25
L318-345 – Although I appreciate these two sections and find them interesting, they are beyond the scope of your study and could thus be reduced to provide more space to discuss your findings.

RESPONSE
We managed to incorporate the reviewer's suggestions for expanding our discussion on the ecological implications of altered behaviours (i.e. comment #24) without needing to reduce these paragraphs to make more space. Hence, we have left these passages as they are.

COMMENT #26
Fig 1: this figure is really appreciated! Perhaps just consider re-organizing the panels following the order of mention in the introduction (panel/scenario 2 is mentioned before scenario 1).

RESPONSE
Thank you. We have switched the position of panel 1 and 2 as suggested.

COMMENT #27
Fig 2: it is a bit complex to make the parallel between your figure 1 and your figure 2. I understand that you may be limited in representing the individual slopes and variations in figure 2 in the same way that you did in figure 1 given your sample size. But perhaps you could add the expected overall repeatability on the three panels of figure 1C to help the reader?
Also, perhaps consider adding letters as you did in figure 1 (a-d) for each panel and cite them in your result section.
Why not including the green to orange gradient shown on figure 1 below the X-axis?
These are just suggestions, as per se your figure is good. Maybe it is just that I put my expectations very high when looking at figure 1, and I would appreciate some visual parallel with your results. But this is me being picky, I admit.

RESPONSE
Done. As already detailed above, we have taken on board these excellent suggestions, which improved the readability of Figure 2 (and the new Figure S2 in the Supplementary Material) and strengthened the link with our predictions in Figure 1C. However, we preferred not to repeat panels from Figure 1C into Figure 2, which has already been expanded to include 2 extra panels and addresses the suggestions from both reviewer 1 (comments #7) and reviewer 2.

REFERENCES

[1] Polverino, G., Cigliano, C., Nakayama, S. & Mehner, T. 2016 Emergence and development of personality over the ontogeny of fish in absence of environmental stress factors. *Behav. Ecol. Sociobiol.* **70**, 2027-2037.
[2] Polverino, G., Palmas, B.M., Evans, J.P. & Gasparini, C. 2019 Individual plasticity in alternative reproductive tactics declines with social experience in male guppies. *Anim. Behav.* **148**, 113-121.
[3] Royauté, R., Buddle, C.M. & Vincent, C. 2015 Under the influence: sublethal exposure to an insecticide affects personality expression in a jumping spider. *Funct. Ecol.* **29**, 962-970.
[4] Geffroy, B., Alfonso, S., Sadoul, B. & Blumstein, D.T. 2020 A world for reactive phenotypes. *Front. Conserv. Sci.* **1**, 611919.

The University of Western Australia
M000, Perth WA 6009 Australia
T +61 8 6488 1473
E giovanni.polverino@uwa.edu.au
CRICOS Provider Code 00126G

[5] Puścian, A., Winiarski, M., Łęski, S., Charzewski, Ł., Nikolaev, T., Borowska, J., Dzik, J.M., Bijata, M., Lipp, H.P. & Dziembowska, M. 2020 Chronic fluoxetine treatment impairs motivation and reward learning by affecting neuronal plasticity in the central amygdala. *Br. J. Pharmacol.* (doi: 10.1111/bph.15319).
[6] McCann, K.S. 2000 The diversity–stability debate. *Nature* **405**, 228.
[7] Bolnick, D.I., Amarasekare, P., Araújo, M.S., Bürger, R., Levine, J.M., Novak, M., Rudolf, V.H.W., Schreiber, S.J., Urban, M.C. & Vasseur, D.A. 2011 Why intraspecific trait variation matters in community ecology. *Trends Ecol. Evol.* **26**, 183-192.
[8] Reid, N.M., Proestou, D.A., Clark, B.W., Warren, W.C., Colbourne, J.K., Shaw, J.R., Karchner, S.I., Hahn, M.E., Nacci, D., Oleksiak, M.F., et al. 2016 The genomic landscape of rapid repeated evolutionary adaptation to toxic pollution in wild fish. *Science* **354**, 1305-1308.
[9] Montiglio, P.O. & Royauté, R. 2014 Contaminants as a neglected source of behavioural variation. *Anim. Behav.* **88**, 29-35.
[10] Saaristo, M., Brodin, T., Balshine, S., Bertram, M.G., Brooks, B.W., Ehlman, S.M., McCallum, E.S., Sih, A., Sundin, J. & Wong, B.B.M. 2018 Direct and indirect effects of chemical contaminants on the behaviour, ecology and evolution of wildlife. *Proc. R. Soc. B.* **285**, 20181297.
[11] Dingemanse, N.J. & Dochtermann, N.A. 2013 Quantifying individual variation in behaviour: mixed-effect modelling approaches. *J. Anim. Ecol.* **82**, 39-54.
[12] Dochtermann, N.A., Schwab, T. & Sih, A. 2015 The contribution of additive genetic variation to personality variation: heritability of personality. *Proc. R. Soc. B.* **282**, 20142201.
[13] Biro, P.A. & Stamps, J.A. 2008 Are animal personality traits linked to life-history productivity? *Trends Ecol. Evol.* **23**, 361-368.
[14] Niemelä, P.T. & Dingemanse, N.J. 2018 Meta-analysis reveals weak associations between intrinsic state and personality. *Proc. R. Soc. B.* **285**, 20172823.
[15] Freund, J., Brandmaier, A.M., Lewejohann, L., Kirste, I., Kritzler, M., Krüger, A., Sachser, N., Lindenberger, U. & Kempermann, G. 2013 Emergence of individuality in genetically identical mice. *Science* **340**, 756-775.
[16] Bell, A.M. 2005 Behavioural differences between individuals and two populations of stickleback (*Gasterosteus aculeatus*). *J. Evol. Biol.* **18**, 464-473.
[17] Bell, A.M. & Sih, A. 2007 Exposure to predation generates personality in threespined sticklebacks (*Gasterosteus aculeatus*). *Ecol. Lett.* **10**, 828-834.
[18] Dingemanse, N.J., Wright, J., Kazem, A.J., Thomas, D.K., Hickling, R. & Dawnay, N. 2007 Behavioural syndromes differ predictably between 12 populations of three-spined stickleback. *J. Anim. Ecol.* **76**, 1128-1138.
[19] Jacquin, L., Petitjean, Q., Côte, J., Laffaille, P. & Jean, S. 2020 Effects of pollution on fish behavior, personality and cognition: some research perspectives. *Front. Ecol. Evol.* **8**, 86.
[20] Wong, D.T., Perry, K.W. & Bymaster, F.P. 2005 The discovery of fluoxetine hydrochloride (Prozac). *Nat. Rev. Drug Discov.* **4**, 764-774.
[21] Bean, T.G., Boxall, A.B.A., Lane, J., Herborn, K.A., Pietravalle, S. & Arnold, K.E. 2014 Behavioural and physiological responses of birds to environmentally relevant concentrations of an antidepressant. *Philos. Trans. R. Soc. Lond. B.* **369**, 20130575.
[22] Tan, H., Polverino, G., Martin, J.M., Bertram, M.G., Wiles, S.C., Palacios, M.M., Bywater, C.L., White, C.R. & Wong, B.B.M. 2020 Chronic exposure to a pervasive pharmaceutical pollutant erodes among-individual phenotypic variation in a fish. *Environ. Pollut.* **263**, 114450.
[23] Houde, A. 2019 *Sex, color, and mate choice in guppies*, Princeton University Press.
[24] Polverino, G., Santostefano, F., Díaz-Gil, C. & Mehner, T. 2018 Ecological conditions drive pace-of-life syndromes by shaping relationships between life history, physiology and behaviour in two populations of Eastern mosquitofish. *Sci. Rep.* **8**, 14673.
[25] Krause, J., Loader, S.P., McDermott, J. & Ruxton, G.D. 1998 Refuge use by fish as a function of body length–related metabolic expenditure and predation risks. *Proc. R. Soc. B.* **265**, 2373-2379.
[26] Bierbach, D., Laskowski, K.L. & Wolf, M. 2017 Behavioural individuality in clonal fish arises despite near-identical rearing conditions. *Nat. Comm.* **8**, 1-7.

The University of Western Australia
M000, Perth WA 6009 Australia
T +61 8 6488 1473
E giovanni.polverino@uwa.edu.au
CRICOS Provider Code 00126G