THE ROYAL SOCIETY PUBLISHING

PROCEEDINGS B

Anthropogenic noise impairs foraging for cryptic prey via cross-sensory interference

Wouter Halfwerk and Kees van Oers

Article citation details
Proc. R. Soc. B **287**: 20192951.
http://dx.doi.org/10.1098/rspb.2019.2951

Review timeline

Original submission:	25 September 2019
1st revised submission:	19 December 2019
2nd revised submission:	6 February 2020
3rd revised submission:	6 March 2020
4th revised submission:	12 March 2020
Final acceptance:	12 March 2020

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

Review History

RSPB-2019-2239.R0 (Original submission)

Review form: Reviewer 1 (Bertrand Lemasson)

Recommendation
Major revision is needed (please make suggestions in comments)

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

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

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

Is the length of the paper justified?
Yes

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

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?
N/A

Is it clear?
N/A

Is it adequate?
N/A

Do you have any ethical concerns with this paper?
No

Comments to the Author
I have attached my comments in a separate word document (See Appendix A).

Review form: Reviewer 2

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?
No

Is it clear?
N/A

Is it adequate?
N/A

Do you have any ethical concerns with this paper?
No

Comments to the Author
How anthropogenic noise affects wildlife is a research topic to have received considerable attention in the last 10-15 years. One issue that has been particularly well-explored is uni-modal interference – i.e. how noise masks acoustic cues/signals or distracts receivers from making appropriate responses to them. By contrast, there has been relatively little testing of cross-modal or cross-sensory effects – i.e. how noise might interfere with the processing of information from, for example, visual or olfactory cues/signals. This paper reports on a series of simple, but elegant, aviary-based experiments with great tits, investigating the effect of additional noise on foraging for prey of varying levels of conspicuousness. The experiments are well-designed and controlled, the analyses appropriate, the findings are novel and of broad appeal, and the paper generally well-written; I enjoyed reading it. I have just some relatively small points which I help will improve still further this interesting work.

Lines 31-43: This is ambiguous. It could imply that the birds were trained during the presence or absence of noise; instead, I think the birds were trained in ambient conditions and then tested in either the presence or absence of noise. So, a little rephrasing here would help.

Line 51: Whilst it is certainly true that initial work in the field of anthropogenic noise impacts on wildlife had a definite focus on masking of acoustic cues and signals, a large number of studies have now explored a range of other behavioural, physiological, developmental and fitness effects arising from the impact of noise as a distractor and/or stressor (see Shannon et al. 2015 Biol. Rev. for what was a comprehensive list of relevant papers at the time).

Lines 53-54: I would also include mammals (marine and terrestrial) in this list. For example, Parks et al. 2010 Biol. Lett provides an example of a change in whale vocalisations in noise, whereas Kern & Radford 2016 Env. Poll. is an example of how terrestrial mammal communication can be negatively affected by noise.

Lines 61-64: First sentence puts the focus on 'non-acoustic' impacts of anthropogenic noise, but second sentence talks about masking within the acoustic channel. Unless I am missing something, these therefore feel at odds with one another.

Lines 84-86: Luo et al. 2015 Glob. Change Biol. and Zhou et al. 2019 Fun. Ecol. provide two examples where there has been a specific consideration of potential mechanisms by which noise can have an impact and so might be worth mentioning as exceptions here.

Lines 101-102: Mention of these two selection lines relating to exploratory behaviour comes rather out of the blue and, indeed, this aspect of the paper is the one that is not really set-up or discussed properly. I would suggest that if it were to remain as part of the results, a little introduction to intraspecific variation in responses to noise in general (see Harding et al. 2019 Behav. Ecol. for a recent review on this topic), and exploratory behaviour / personality specifically (see Naguib et al. 2013 Anim. Behav. for a personality-related effect of noise), would be useful (as well as a suitable prediction about what might be expected with respect to this difference).

Lines 110-117: It is not completely clear here how the second two experiments differ from one another; a little extra clarity would be useful.

Line 131: What was the split in terms of sexes and selection lines among the 17 birds that were successfully trained?

Lines 154-155: Was there a set / formal criteria here (e.g. 10 approaches out of 12)?

Lines 177-180: How many exemplars of each type were created? Different ones for every trial?

Lines 186-188: How many different noise files were created and used?

Lines 204-205: Slight ambiguity as it sounds as if noise levels were gradually increased in control trials, which seems unlikely as those entailed the playback of silence.

Lines 210-214: How many trials were abandoned for each of these reasons?

Lines 231-232: Why was this additional training round introduced?

Line 235: What was the sex and selection line split of these remaining 12 birds?

Lines 243-244: It is not completely clear what this experiment entailed, in terms of what was presented during a particular trial and what was presented on day one vs day two. Also, why the change to the presentation of one target instead of two on day two?

Lines 253-255: Why was there scoring by the experimenter if there was a blind scoring option available; the latter is surely the better bet?

Line 264: What % matched (more useful than saying "mostly")?

Line 265: How many cases?

Lines 278-280: Given there is mention of the number of males and females earlier in the Methods, why was sex not included as a potential fixed factor?

Lines 352-354: These sentences confused me on first reading because the final section of the Results is the only part to mention habituation explicitly… and there the emphasis is on a lack of habituation, which is also the message from Figure 1D. So, a bit of extra clarity here would be useful.

Line 360: It feels like it would be useful to include a few sentences on how the current findings relate to previous work examining foraging and cross-modal/sensory impacts of noise on wildlife; the Discussion is currently a little light on information relating to the main results and relevant literature.

Lines 390-393: As with the Introduction, if the comparison of selection lines is to be retained then it needs to be given a little more airtime in the paper – here, that could be how the finding relates to previous work on intraspecific variation in noise responses.

Lines 403-406: There are now several studies that have shown differences in responses of animals to noise dependent on previous experience/exposure, which are directly relevant here (see Harding et al. 2019 Behav. Ecol. for table of relevant examples).

Figure 2: Given the first sentence of the figure caption compares across the three panels, it might be useful if those included mean values for ease of reader comparison.

Figure 3: It might be worth considering including both noise and quiet values for each of the three moth types, thus showcasing the lack of sound-treatment effect on the probability to attack.

Decision letter (RSPB-2019-2239.R0)

06-Nov-2019

Dear Dr Halfwerk:

I am writing to inform you that your manuscript RSPB-2019-2239 entitled "Anthropogenic noise impairs foraging for cryptic prey via cross-sensory interference" has, in its current form, been rejected for publication in Proceedings B.

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

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

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

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

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

Sincerely,
Dr Daniel Costa
mailto: proceedingsb@royalsociety.org

Associate Editor
Board Member: 1
Comments to Author:
Two expert reviewers have now assessed your manuscript and both are quite positive (particularly reviewer 2). However both have a number of minor (reviewer 2) and major (reviewer 1) comments and criticisms that need to be addressed prior to publication. In particular, reviewer 2 found sections of the manuscript confusing (particularly the methods and results) and has detailed suggestions for improvement. I am convinced that a careful re-writing

and/or clarification of the various sections of text spotlighted by the reviewer will fix the problem.

Reviewer(s)' Comments to Author:

Referee: 1
Comments to the Author(s)
I have attached my comments in a separate word document.

Referee: 2
Comments to the Author(s)
How anthropogenic noise affects wildlife is a research topic to have received considerable attention in the last 10-15 years. One issue that has been particularly well-explored is uni-modal interference – i.e. how noise masks acoustic cues/signals or distracts receivers from making appropriate responses to them. By contrast, there has been relatively little testing of cross-modal or cross-sensory effects – i.e. how noise might interfere with the processing of information from, for example, visual or olfactory cues/signals. This paper reports on a series of simple, but elegant, aviary-based experiments with great tits, investigating the effect of additional noise on foraging for prey of varying levels of conspicuousness. The experiments are well-designed and controlled, the analyses appropriate, the findings are novel and of broad appeal, and the paper generally well-written; I enjoyed reading it. I have just some relatively small points which I help will improve still further this interesting work.

Lines 31-43: This is ambiguous. It could imply that the birds were trained during the presence or absence of noise; instead, I think the birds were trained in ambient conditions and then tested in either the presence or absence of noise. So, a little rephrasing here would help.

Line 51: Whilst it is certainly true that initial work in the field of anthropogenic noise impacts on wildlife had a definite focus on masking of acoustic cues and signals, a large number of studies have now explored a range of other behavioural, physiological, developmental and fitness effects arising from the impact of noise as a distractor and/or stressor (see Shannon et al. 2015 Biol. Rev. for what was a comprehensive list of relevant papers at the time).

Lines 53-54: I would also include mammals (marine and terrestrial) in this list. For example, Parks et al. 2010 Biol. Lett provides an example of a change in whale vocalisations in noise, whereas Kern & Radford 2016 Env. Poll. is an example of how terrestrial mammal communication can be negatively affected by noise.

Lines 61-64: First sentence puts the focus on 'non-acoustic' impacts of anthropogenic noise, but second sentence talks about masking within the acoustic channel. Unless I am missing something, these therefore feel at odds with one another.

Lines 84-86: Luo et al. 2015 Glob. Change Biol. and Zhou et al. 2019 Fun. Ecol. provide two examples where there has been a specific consideration of potential mechanisms by which noise can have an impact and so might be worth mentioning as exceptions here.

Lines 101-102: Mention of these two selection lines relating to exploratory behaviour comes rather out of the blue and, indeed, this aspect of the paper is the one that is not really set-up or discussed properly. I would suggest that if it were to remain as part of the results, a little introduction to intraspecific variation in responses to noise in general (see Harding et al. 2019 Behav. Ecol. for a recent review on this topic), and exploratory behaviour / personality specifically (see Naguib et al. 2013 Anim. Behav. for a personality-related effect of noise), would be useful (as well as a suitable prediction about what might be expected with respect to this difference).

Lines 110-117: It is not completely clear here how the second two experiments differ from one another; a little extra clarity would be useful.

Line 131: What was the split in terms of sexes and selection lines among the 17 birds that were successfully trained?

Lines 154-155: Was there a set / formal criteria here (e.g. 10 approaches out of 12)?

Lines 177-180: How many exemplars of each type were created? Different ones for every trial?

Lines 186-188: How many different noise files were created and used?

Lines 204-205: Slight ambiguity as it sounds as if noise levels were gradually increased in control trials, which seems unlikely as those entailed the playback of silence.

Lines 210-214: How many trials were abandoned for each of these reasons?

Lines 231-232: Why was this additional training round introduced?

Line 235: What was the sex and selection line split of these remaining 12 birds?

Lines 243-244: It is not completely clear what this experiment entailed, in terms of what was presented during a particular trial and what was presented on day one vs day two. Also, why the change to the presentation of one target instead of two on day two?

Lines 253-255: Why was there scoring by the experimenter if there was a blind scoring option available; the latter is surely the better bet?

Line 264: What % matched (more useful than saying "mostly")?

Line 265: How many cases?

Lines 278-280: Given there is mention of the number of males and females earlier in the Methods, why was sex not included as a potential fixed factor?

Lines 352-354: These sentences confused me on first reading because the final section of the Results is the only part to mention habituation explicitly… and there the emphasis is on a lack of habituation, which is also the message from Figure 1D. So, a bit of extra clarity here would be useful.

Line 360: It feels like it would be useful to include a few sentences on how the current findings relate to previous work examining foraging and cross-modal/sensory impacts of noise on wildlife; the Discussion is currently a little light on information relating to the main results and relevant literature.

Lines 390-393: As with the Introduction, if the comparison of selection lines is to be retained then it needs to be given a little more airtime in the paper – here, that could be how the finding relates to previous work on intraspecific variation in noise responses.

Lines 403-406: There are now several studies that have shown differences in responses of animals to noise dependent on previous experience/exposure, which are directly relevant here (see Harding et al. 2019 Behav. Ecol. for table of relevant examples).

Figure 2: Given the first sentence of the figure caption compares across the three panels, it might be useful if those included mean values for ease of reader comparison.

Figure 3: It might be worth considering including both noise and quiet values for each of the three moth types, thus showcasing the lack of sound-treatment effect on the probability to attack.

Author's Response to Decision Letter for (RSPB-2019-2239.R0)

See Appendix B.

RSPB-2019-2951.R0

Review form: Reviewer 1

Recommendation
Major revision is needed (please make suggestions in comments)

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

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

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

Is the length of the paper justified?
Yes

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

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?
N/A

Is it clear?
N/A

Is it adequate?
N/A

Do you have any ethical concerns with this paper?
No

Comments to the Author

In general, I have enjoyed reading this study and I found that the authors addressed most of the points brought up by both reviewers and I have only a few minor lingering replies. However, my concerns over the analysis and interpretation of experiment 3 remain. The third experiment is effectively about testing for evidence of cross-modal sensory disruption that impacts a higher-order cognitive process, making a decision – an intention that certainly increases the impact of the study. Based on my interpretation of the descriptions provided, I do not agree with some of the assumptions and so I find myself questioning the results. The description of the new methodology for this experiment remains unnecessarily confusing, as details of the statistical approach vary between sections of the manuscript. I believe that the authors should re-evaluate their approach regarding experiment three and provide some biological arguments for how they choose to proceed. Below I provide details to explain my position and hopefully provide useful feedback.

Minor comments

AIC - You state that you use AIC corrected for small values. Do you mean AICc? I ask because you report AIC everywhere.

Overfitting - AIC assesses relative fit among competing models, but it does not address the question of overall fit. In fact, AIC is biased towards selecting overfit models. You cite Zuur et al. 2009 as one of your references. I think you'll find that Zuur generally encourages simulating ones' final model to assess fit. If you have an overfit model, your simulations will show it because the (original) overfit model was mostly fitting the noise rather than the general trends in your data. The glmmTMB package has made the simulation of glmm's pretty straightforward.

Pseudo-replication - On line 220 you state that 8 different oak trees were used as sources for your target stimuli and background prints to 'avoid pseudo-replication'. I don't see the point of this statement since any concerns over pseudo-replication apply to the repeated measurements on the same birds (within subject) not on whether or not they were tested on the same tree background. In any case, you account for this with your random effects.

Major comments: experiment 3

I don't believe you are correctly testing your hypothesis here and so your statistics do not support your conclusions. Discerning this was more effort than it should be. In the methods [l. 307-312] you describe a model to test your hypothesis, while in the SI [SI, Exp. 3] you present a different model, and your wording in the results [l. 364-370] suggests that you actually applied both models separately. If that is the case, I do not agree with this approach for two reasons: i) it ignores any interactive effect between each of your targets on bird choice and ii) your noise treatment seems unbalanced.

The hypothesis being tested here [l. 114-116] is that noise will influence a bird's choice when faced with targets differing in their level of visual crypsis. Ignoring the noise treatment for a moment, an individual's choice is not only dependent on the level of a given target's crypsis, but also on the opposing target's level of crypsis (you state this yourselves in one of your responses). The model you describe in the methods [l. 308-312] suggests that you actually did something like Y1 ~ T1 x NoiseTreatment + T2 + (Rand. Eff) + GlobalError, where Y1 is the attack score on target 1, T1 and T2 are the corresponding characteristics of the targets (white, colour, or cryptic), and noise is the presence/absence of your sound treatment (+ your random effect and global error terms). If you simply add target 2 as another fixed effect (I think the term covariate is incorrect here), then you are assuming that both targets are independently influencing a given bird's decision to attack target 1. This is equivalent to stating that the chances of attacking a colour target is independent of there being either a white one or a cryptic one on the other 'tree' - an assumption that strikes me as incorrect. You need an interaction between targets. All this being said, you may reach the same conclusions if the interaction term is not significant, but the targets

themselves each are. The difference would be that you didn't explicitly ignore the interactive potential, but tested for it and found no evidence of it.

Given the above, how you model the noise should be reconsidered carefully and defended. Following precedence from experiments 1 & 2, one may expect you to include an interactive effect of noise just as you did with latency (in fact, this is what you were doing, albeit incorrectly in my opinion). However, this may be problematic and it's not apparent to me whether or not it is the correct approach. It would, however, be the most parsimonious approach since, by your own words (397-399) don't know how noise is disrupting the decision process. Unfortunately, by accounting for the potential interaction between targets you would be creating a 3-way interaction between each target and the noise, which can be challenging to fit and difficult to interpret. If this approach works, however, it would provide some insight to your point on lines 397-399 in the discussion.

Your approach with latency in Exp. 3 suffers from the same issues outlined above for the analysis of target choice. However, not having to specify a particular target as your response variable would seem to allow for additional latitude on how you approach this problem. For instance, if a 3-way interaction proves to be too much for your sample size, couldn't you use the stimulus pairing treatment as an index of relative difficulty (supported by your results in Exp. 2)? If so, you might consider simply regressing latency on the interaction between stimulus treatment x noise.

Decision letter (RSPB-2019-2951.R0)

21-Jan-2020

Dear Dr Halfwerk:

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 still have some concerns with your manuscript. We normally do not allow multiple rounds of revision, but I am willing to allow you one final opportunity to fully respond and revise your manuscript.

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. 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
Comments to Author:
One reviewer has now appraised your re-submitted manuscript, and while they are generally very positive, they have difficulties with the assumptions, design and interpretation of Experiment 3. The reviewer feels that you are not actually testing your hypothesis in the correct manner, and that the models chosen to test the hypothesis may not be appropriate. Thus, a careful reconsideration of these issues and an appropriate revision will be necessary prior to publication.

Reviewer(s)' Comments to Author:

Referee: 1

Comments to the Author(s).
In general, I have enjoyed reading this study and I found that the authors addressed most of the points brought up by both reviewers and I have only a few minor lingering replies. However, my concerns over the analysis and interpretation of experiment 3 remain. The third experiment is effectively about testing for evidence of cross-modal sensory disruption that impacts a higher-order cognitive process, making a decision – an intention that certainly increases the impact of the study. Based on my interpretation of the descriptions provided, I do not agree with some of the assumptions and so I find myself questioning the results. The description of the new methodology for this experiment remains unnecessarily confusing, as details of the statistical approach vary between sections of the manuscript. I believe that the authors should re-evaluate their approach regarding experiment three and provide some biological arguments for how they choose to proceed. Below I provide details to explain my position and hopefully provide useful feedback.

Minor comments

AIC - You state that you use AIC corrected for small values. Do you mean AICc? I ask because you report AIC everywhere.

Overfitting - AIC assesses relative fit among competing models, but it does not address the question of overall fit. In fact, AIC is biased towards selecting overfit models. You cite Zuur et al. 2009 as one of your references. I think you'll find that Zuur generally encourages simulating ones' final model to assess fit. If you have an overfit model, your simulations will show it because the (original) overfit model was mostly fitting the noise rather than the general trends in your data. The glmmTMB package has made the simulation of glmm's pretty straightforward.

Pseudo-replication - On line 220 you state that 8 different oak trees were used as sources for your target stimuli and background prints to 'avoid pseudo-replication'. I don't see the point of this statement since any concerns over pseudo-replication apply to the repeated measurements on the same birds (within subject) not on whether or not they were tested on the same tree background. In any case, you account for this with your random effects.

Major comments: experiment 3

I don't believe you are correctly testing your hypothesis here and so your statistics do not support

your conclusions. Discerning this was more effort than it should be. In the methods [l. 307-312] you describe a model to test your hypothesis, while in the SI [SI, Exp. 3] you present a different model, and your wording in the results [l. 364-370] suggests that you actually applied both models separately. If that is the case, I do not agree with this approach for two reasons: i) it ignores any interactive effect between each of your targets on bird choice and ii) your noise treatment seems unbalanced.

The hypothesis being tested here [l. 114-116] is that noise will influence a bird's choice when faced with targets differing in their level of visual crypsis. Ignoring the noise treatment for a moment, an individual's choice is not only dependent on the level of a given target's crypsis, but also on the opposing target's level of crypsis (you state this yourselves in one of your responses). The model you describe in the methods [l. 308-312] suggests that you actually did something like Y1 ~ T1 x NoiseTreatment + T2 + (Rand. Eff) + GlobalError, where Y1 is the attack score on target 1, T1 and T2 are the corresponding characteristics of the targets (white, colour, or cryptic), and noise is the presence/absence of your sound treatment (+ your random effect and global error terms). If you simply add target 2 as another fixed effect (I think the term covariate is incorrect here), then you are assuming that both targets are independently influencing a given bird's decision to attack target 1. This is equivalent to stating that the chances of attacking a colour target is independent of there being either a white one or a cryptic one on the other 'tree' - an assumption that strikes me as incorrect. You need an interaction between targets. All this being said, you may reach the same conclusions if the interaction term is not significant, but the targets themselves each are. The difference would be that you didn't explicitly ignore the interactive potential, but tested for it and found no evidence of it.

Given the above, how you model the noise should be reconsidered carefully and defended. Following precedence from experiments 1 & 2, one may expect you to include an interactive effect of noise just as you did with latency (in fact, this is what you were doing, albeit incorrectly in my opinion). However, this may be problematic and it's not apparent to me whether or not it is the correct approach. It would, however, be the most parsimonious approach since, by your own words (397-399) don't know how noise is disrupting the decision process. Unfortunately, by accounting for the potential interaction between targets you would be creating a 3-way interaction between each target and the noise, which can be challenging to fit and difficult to interpret. If this approach works, however, it would provide some insight to your point on lines 397-399 in the discussion.

Your approach with latency in Exp. 3 suffers from the same issues outlined above for the analysis of target choice. However, not having to specify a particular target as your response variable would seem to allow for additional latitude on how you approach this problem. For instance, if a 3-way interaction proves to be too much for your sample size, couldn't you use the stimulus pairing treatment as an index of relative difficulty (supported by your results in Exp. 2)? If so, you might consider simply regressing latency on the interaction between stimulus treatment x noise.

Author's Response to Decision Letter for (RSPB-2019-2951.R0)

See Appendix C.

RSPB-2019-2951.R1 (Revision)

Review form: Reviewer 1

Recommendation
Major revision is needed (please make suggestions in comments)

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

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

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

Is the length of the paper justified?
Yes

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

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?
N/A

Is it clear?
N/A

Is it adequate?
N/A

Do you have any ethical concerns with this paper?
No

Comments to the Author
In reading this 3rd version of the paper by Halfwerk & van Oers I will admit that I am torn, but I must suggest that the manuscript not be accepted in its current form. I believe that the first two experiments provide a valuable contribution to the field and I agree with the authors that their main conclusions rest on the findings from these data. So, on the one hand the paper would not require major revisions to be acceptable. On the other hand, I find the authors' rational and methodology for the third experiment unconvincing. In my opinion, a viable strategy moving forward would be to reassess the reasons for, and analysis of, the final experiment or to drop it entirely.

The authors now argue that the last experiment had two objectives. The first being a need to verify that the birds were using visual cues and not a random strategy and the second to test for

the effects of noise on target choice. The authors have already controlled for the random search strategy in their previous experiments by first randomizing the placement of the targets and by the design of experiment 2. (As an aside, in the original version the statement about randomizing the target's location among the pegs was phrased in a manner that suggested that this was done for all trials, yet in this version the text reads as if this were only done for experiments 1 & 2 – an odd discrepancy that now substantiates the need for experiment 3, but is not highlighted as a modification.) The authors also noted that the second round of experiment 2 definitively showed that all subjects had been responding to the visual cue (l. 353-355), which also seems to argue against any concerns of a random search strategy.

Regarding the second reason provided for experiment 3, noise effects on target choice, the analysis still makes little sense to me (I have no problem with latency). I'm not familiar with a case in which a GLMM is used to predict a categorical dependent variable having three possible options and I'm having trouble following the logic. What exactly is being predicted? A Poisson distribution deals with the probability of seeing Y (an integer value) with a mean value of mu. If you have three possible targets that could be presented as a choice, what does the mean represent? Would this value represent the mean number of whites, color, or cryptic? I would also note that the alternative approach described in the response to my comments (of modeling the choice to attack one of the two available targets as a binomial function of that target's crypsis level) ignores the presence of the alternative choice and therefore distills down to a repeat of experiment 2.

Decision letter (RSPB-2019-2951.R1)

24-Feb-2020

Dear Dr Halfwerk:

Your manuscript has now been peer reviewed. The reviewers' comments (not including confidential comments to the Editor) are included at the end of this email for your reference. As you will see, the reviewer still has some issues with with your manuscript and we would like to invite you to respond and revise your manuscript to address them.

This will be your final opportunity to address these issues as we do not normally allow multiple rounds of revision. If deemed necessary by the Associate Editor, your manuscript will be sent back to a different reviewer for assessment. 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. 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
Editor, Proceedings B
mailto: proceedingsb@royalsociety.org

Reviewer(s)' Comments to Author:

Referee: 1
Comments to the Author(s)
In reading this 3rd version of the paper by Halfwerk & van Oers I will admit that I am torn, but I must suggest that the manuscript not be accepted in its current form. I believe that the first two experiments provide a valuable contribution to the field and I agree with the authors that their main conclusions rest on the findings from these data. So, on the one hand the paper would not require major revisions to be acceptable. On the other hand, I find the authors' rational and methodology for the third experiment unconvincing. In my opinion, a viable strategy moving forward would be to reassess the reasons for, and analysis of, the final experiment or to drop it entirely.

The authors now argue that the last experiment had two objectives. The first being a need to verify that the birds were using visual cues and not a random strategy and the second to test for the effects of noise on target choice. The authors have already controlled for the random search strategy in their previous experiments by first randomizing the placement of the targets and by the design of experiment 2. (As an aside, in the original version the statement about randomizing the target's location among the pegs was phrased in a manner that suggested that this was done for all trials, yet in this version the text reads as if this were only done for experiments 1 & 2 – an odd discrepancy that now substantiates the need for experiment 3, but is not highlighted as a modification.) The authors also noted that the second round of experiment 2 definitively showed that all subjects had been responding to the visual cue (l. 353-355), which also seems to argue against any concerns of a random search strategy.

Regarding the second reason provided for experiment 3, noise effects on target choice, the analysis still makes little sense to me (I have no problem with latency). I'm not familiar with a case in which a GLMM is used to predict a categorical dependent variable having three possible options and I'm having trouble following the logic. What exactly is being predicted? A Poisson distribution deals with the probability of seeing Y (an integer value) with a mean value of mu. If you have three possible targets that could be presented as a choice, what does the mean represent? Would this value represent the mean number of whites, color, or cryptic? I would also note that the alternative approach described in the response to my comments (of modeling the choice to attack one of the two available targets as a binomial function of that target's crypsis level) ignores the presence of the alternative choice and therefore distills down to a repeat of experiment 2.

Decision letter (RSPB-2019-2951.R2)

12-Mar-2020

Dear Dr Halfwerk

I am pleased to inform you that your Review manuscript RSPB-2019-2951.R2 entitled "Anthropogenic noise impairs foraging for cryptic prey via cross-sensory interference" has been accepted for publication in Proceedings B.

The referee(s) do not recommend any further changes. Therefore, please proof-read your manuscript carefully and upload your final files for publication. Because the schedule for publication is very tight, it is a condition of publication that you submit the revised version of your manuscript within 7 days. If you do not think you will be able to meet this date please let me know immediately.

To upload your manuscript, log into http://mc.manuscriptcentral.com/prsb and enter your Author Centre, where you will find your manuscript title listed under "Manuscripts with Decisions." Under "Actions," click on "Create a Revision." Your manuscript number has been appended to denote a revision.

You will be unable to make your revisions on the originally submitted version of the manuscript. Instead, upload a new version through your Author Centre.

Before uploading your revised files please make sure that you have:
1) A text file of the manuscript (doc, txt, rtf or tex), including the references, tables (including captions) and figure captions. Please remove any tracked changes from the text before submission. PDF files are not an accepted format for the "Main Document".

2) A separate electronic file of each figure (tiff, EPS or print-quality PDF preferred). The format should be produced directly from original creation package, or original software format. Please note that PowerPoint files are not accepted.

3) Electronic supplementary material: this should be contained in a separate file from the main text and the file name should contain the author's name and journal name, e.g authorname_procb_ESM_figures.pdf
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 see: https://royalsociety.org/journals/authors/author-guidelines/

4) Data-Sharing and data citation
It is a condition of publication that data supporting your paper are made available. Data should be made available either in the electronic supplementary material or through an appropriate repository. Details of how to access data should be included in your paper. Please see https://royalsociety.org/journals/ethics-policies/data-sharing-mining/ for more details.

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=RSPB-2019-2951.R2 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.

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

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

Sincerely,
Dr Daniel Costa
Editor, Proceedings B

mailto:proceedingsb@royalsociety.org

Associate Editor Board Member
Comments to Author:
Thank you for taking the effort to consider the sound recommendations of the final reviewer by removing the third experiment. I feel that the manuscript is now ready for publication, and that your study is an important contribution to the field. I have only one exceptionally small suggestion for improvement: in Figure 1, the label "D" for panel D isn't actually sitting over that panel. Can you move it to the right so that it is?

Decision letter (RSPB-2019-2951.R3)

12-Mar-2020

Dear Dr Halfwerk

I am pleased to inform you that your manuscript entitled "Anthropogenic noise impairs foraging for cryptic prey via cross-sensory interference" 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

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

Open Access
You are invited to opt for Open Access, making your freely available to all as soon as it is ready for publication under a CCBY licence. Our article processing charge for Open Access is £1700. Corresponding authors from member institutions (http://royalsocietypublishing.org/site/librarians/allmembers.xhtml) receive a 25% discount to these charges. For more information please visit http://royalsocietypublishing.org/open-access.

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

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

You are allowed to post any version of your manuscript on a personal website, repository or preprint server. However, the work remains under media embargo and you should not discuss it

with the press until the date of publication. Please visit https://royalsociety.org/journals/ethics-policies/media-embargo for more information.

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

Sincerely,
Editor, Proceedings B
mailto: proceedingsb@royalsociety.org

Appendix A

Proc Roy. Soc. B rev. notes

In this manuscript by Halfwerk and van Oers the authors report their efforts to determine if noise in one sensory domain can affect individual performance in another. Specifically, the investigators tested if different levels of auditory noise affected an insectivorous bird's ability to capture prey that varied in their degree of visual camouflage. The authors used a line of captive Great tits descended from two populations selected for their exploratory behavior (fast *vs.* slow, which I equated with bold *vs.* timid behaviors). The hypothesis that acoustic noise can interfere with other sensory modalities is an interesting question and certainly falls within the scope of this journal. In practice, however, I do not believe that the manuscript is ready for publication. I found the description of the methods and results a bit confusing. I believe that these concerns may be alleviated through better communication and presentation of the data. Below I list some of my concerns and I hope that the authors find my suggestions helpful.

- *Language* – (minor) In general the writing is fair, with only minor grammatical errors here and there (e.g., misspellings, improper tenses). However, the cumulative effect of these minor errors compounds the larger problems the reader faces when trying to understand the methods and results.

- *Genetic lineages* – The authors don't clarify at the outset why two different populations of birds are tested. As such, the reader is left with the assumption that this was done merely because they were available, rather than to address a specific question. Perhaps move some of the points raised in the discussion regarding this factor to the introduction.

- *Methods*
    - [l., 152] The reason behind using black and white colors is never made clear.

    - [l., 154] The point at which training ends and the experiments begin seems arbitrary. Be precise.

    - [l., 185] define pink noise for the reader

    - [l., 205] I found the application of the noise treatment here a bit confusing, given that you previously stated (l., 197) that subjects were acclimatized to the noise treatment while still in their cages. How then, does each trial start with 30s of noise playback that gradually increases?

    - [l., 208] Explain why you defined an 'attack' the way you did. Presumably, the birds usually did not try to eat the target but merely inspected it.

    - *Background manipulation* – In general, the presentation of two different tree images with different backgrounds way an important source of confusion for me in these experiments. Line 215 states that 2 printed trees are used during the experiments, so presumably you mean all three experiments. Why present the birds with two trees and why must they have different backgrounds? Given that crypsis is clearly a function of the color and patterning of an object relative to its background, the background is clearly an important component. Yet, the reader never gets a sense for how the authors are varying the backgrounds, aside from stating that you

presented 2 of 8 possible images during the experiments. Are you randomly selecting two during each trial or are you just using the same two across all trials (the language suggests the latter)?

Importantly, there's no discussion of how changes in the background influence the response metric. For example, when you use colored or cryptic models (experiments 2 and 3) are these presented equally on both types of tree images? Do you randomize the positions of these background images or are they always on the same side? It is not apparent that you controlled for this added source of noise in your analyses. Even if you match the color target and the cryptic target to their particular tree within a trial, are you balancing the matching of these targets on each tree image? Given that tree background is neither a fixed nor a random effect in your models, it's not clear to me how you account for this factor. How do you know that the background itself had no effect on their behavior?

If two trees are always being presented, why not provide an analysis on individual choice in each experiment, rather than just in the last experiment? At a minimum, a binomial test would indicate whether the birds explored the tree image containing the target more often than chance could explain. In addition, the presence of two trees raises the question as to what data are used to calculate latencies. Are you only analyzing data from trials in which the birds selected the tree with the target? This seems obvious for attack latencies, but what about approach latencies? Please be more specific.

- [l.227] I'd consider 17 subjects a modest, not large, number of subjects. Also, what are 'exploration scores'? Do you mean your approach and attack latencies you describe later?

- [l. 236] Why is this particular bird's behavior problematic? If you randomize the location of the target each time (l., 221), shouldn't that control for how long they take to 'attack' a prey item once they've approached the board? Unless data from this subject comes out as an influential outlier during your model validation phase, I don't see why you'd discard this bird's data.

- [l. 241] – The end of this sentence is confusing '…but balanced across individuals'. Couldn't you just state that you ran a balanced full factorial design?

- [l. 246] In general, I feel that you can do a better job of presenting the rational and methods for this last experiment. In the introduction you state that the reason for providing subjects with choices of varying difficulty was to 'ensure that birds relied on visual information during their attack and approach behavior' (l., 116). Yet, the individual and interactive effects of auditory noise and visual information seem to be components of experiment two, not three. Upon reading the description of experiment three on line 246, and your interpretation of these results in the discussion, it seems that this experiment is more about how auditory noise can influence choice between targets whose relative visibility varies - a harder task and one worth pursuing, but not clearly explained.

- *Statistical analyses* - I have some concerns with the validity of the tests conducted that could use some clarification by the authors. Basically, you are fitting complicated hierarchical models to small data sets and this raises some concerns over the validity of your significance tests. I admit that I have much to learn about hierarchical modeling, but I do feel that my experience with mixed-models is fairly representative of ethologists/ecologists in general.

    - You state that "various random structures" were fitted [l. 270], but you only provide one. You have also chosen an interesting way to account for individual variation by using perch as your random intercept and subject id as the random slope. I'm not sure that this is the wrong way to model your data, but it is a bit odd and unconventional, so it merits more explanation than you have provided.

      For example, while random slope models are uncommon in ecology, I agree that there is a logic to including both a random effect for the intercept and slope here. However, I fear that you are overparameterizing your models given your sample size (which is modest to begin with and drops over the course of the experiments). Random slope models are known to require a large number of data points to avoid spurious results (See Harrison et al. 2018. PeerJ., doi: 10.7717/peerj.4794 for a recent review). In experiment 2 you have just 9 birds, yet you have a model with two categorical predictors (with 2 and 3 levels, respectively), an interaction between these factors, and two random effects. This model strikes me as running the risk of being overparameterized and providing spurious results. I strongly suggest that the authors provide some supplemental material where the reader can access details supporting the validity of their approach without being distracting from the flow of the paper. Alternatively, you could simply use bird id as a random intercept, which would somewhat simplify the models and account for inherent differences in average performance and issues of independence.

      If you wish to retain a random slope, why is perch a random effect as oppose to selection line? I can see how initial perch position could impact a random search, but since you randomized the position of the target across trials then you have already controlled for this nuisance. It seems more biologically intuitive to nest individual id within selection line rather than within initial perch position during approach (Fig. C seems to reinforce nesting id within lineage). Given that your subjects are drawn from two different populations selected for exploratory/bold behavior, one could reasonably expect *a priori* to see differences in latency. Within these populations, we could then also expect effect strength to vary by individual.

    - [l. 280] – If you are still using birds from two selection lines, and your results from experiment 1 show that selection line influences approach latency, why is this term dropped from your model here? I don't see how showing that both selection lines are influenced by noise tells you anything about how they impact individual performance in the visual task.

    - [l. 281] While I appreciate the intent of experiment three, I found the description of your analysis here rather confusing. Why is noise the only predictor of latency here, but both noise and target combinations are used to predict choice? How is choice scored here anyway? If

you're assuming binomial errors, doesn't that mean that your response is either a 0 or 1? If so, I'm a bit confused by how you score the results from a given trial in which two different targets are presented. On line 275 you state that you scored choice *vs.* no choice. Since you only present an analysis of choice in your third experiment, I'm not sure what this means.

A (perhaps) minor point here that relates to the above. Line 276 you refer to using the cbind function to compare choice. Do you mean the function in R to combine values of vectors, columns, or matrices? If so, I'm not sure how your use of this tool clarifies how you scored your response variable for experiment three.

- [l. 284] The number of birds keeps changing between experiments. Please provide an exact count for each experiment when you introduce them, rather than making the reader go back and calculate how many birds we are down to now or wait for the statistics.

- Results

  - Fig 1c – why show only the approach latency data for this combination of fixed factors? Even if you found no effect of selection line on attack latencies, the reader is left wondering what those responses looked like. Also, the figure caption says that the boxplots are of the predicted values (model estimates). Is that right, or do you mean the actual data?

  - [l. 304] 'comparable numbers from both selection lines' is too vague. Be precise.

  - [l. 304-308] I found this confusing. Are days 2 & 3 in figure 1D showing results during training runs? Why?

  - Figure 2 – I didn't find this figure particularly informative. The text is primarily reporting average differences in approach and attack latencies among levels of crypsis. This information is lost in figure 2, which instead focuses on showing individual-level variation rather than group level differences. Also, why do you again not show results for both approach and attack latency data (particularly since both are apparently influential here)?

  - Figure 3 is referenced when presenting the results of the third experiment, but this figure did little to help reduce my confusion. For example, [l., 325] 'a post-hoc analysis revealed that birds preferred to attack the conspicuous prey over the cryptic prey (white *vs.* cryptic, … Fig 3). White *vs.* cryptic is a level in your fixed effect, so if this combination influenced choice then what was the choice (white or cryptic)? Figure 3 shows the proportion of times that each level of crypsis was selected but doesn't convey how these number vary due to the choice presented. That is, you are not merely asking what the probability of selecting a white target is, but what is the probability of selecting a white target when a colored one is present or when a cryptic one is present? If there is a nearly 60% chance that a bird attacked a white target, how is that number affected by whether the white target was presented with a colored one or the cryptic one? Again, the methods and analysis of this experiment need to be clarified.

- [l. 335- 340, Fig. 1D] – While it is worth pointing out to the reader that auditory noise consistently impacted both approach and attack latencies across these experiments, I don't see how you can pool the data from these different experiments and tack on a mysterious P-value. This seems ill-advised and devalues what is otherwise a perfectly legitimate observation.

- Discussion

  - [l., 351] 'noise reduced foraging for conspicuous prey on the first trial, but not subsequent trials'. It's unclear what you are referring to here.

  - [l., 382] – This statement raises an interesting point and you seem to offer an explanation based on this concept of habituation. However, in going back through your results I don't find any mention of any analysis to support this statement *vs.* getting these results by chance.

  - [l. 387] Given my confusion regarding your analysis of experiment three, this statement about higher-level cognitive effects is not convincing.

Appendix B

Dear editor,

We are grateful for the comments by both reviewers and for the opportunity to resubmit our manuscript. We have provided additional information to the experimental design and statistics in response to some of the comments of reviewer 1. We have also reran several models in response to his/her comments and report the new outputs from these models. Finally, we have added a supplementary file where we reported the outcome of our model selection procedure as well as the output of the full model. These new analyses did not alter our main conclusions but did require some changes to the results section and the figures.

Below we provide a detailed reply to each comment separately. We think that we have addressed all the comments appropriately and we hope that is now suitable for publication.

Kind regards,

Wouter Halfwerk & Kees van Oers

Comments to the reviewers

Associate Editor
Board Member: 1
Comments to Author:
Two expert reviewers have now assessed your manuscript and both are quite positive (particularly reviewer 2). However both have a number of minor (reviewer 2) and major (reviewer 1) comments and criticisms that need to be addressed prior to publication. In particular, reviewer 2 found sections of the manuscript confusing (particularly the methods and results) and has detailed suggestions for improvement. I am convinced that a careful re-writing and/or clarification of the various sections of text spotlighted by the reviewer will fix the problem.

Reviewer(s)' Comments to Author:

Referee: 1

We have pasted the referee's comments and our replies (in red) below.

In this manuscript by Halfwerk and van Oers the authors report their efforts to determine if noise in one sensory domain can affect individual performance in another. Specifically, the investigators tested if different levels of auditory noise affected an insectivorous bird's ability to capture prey that varied in their degree of visual camouflage. The authors used a line of captive Great tits descended from two populations selected for their exploratory behavior (fast *vs.* slow, which I equated with bold *vs.* timid behaviors). The hypothesis that acoustic noise can interfere with other sensory modalities is an interesting question and certainly falls within the scope of this journal. In practice, however, I do not believe that the manuscript is ready for publication. I found the description of the methods and results a bit confusing. I believe that these concerns may be alleviated through better communication and presentation of the data. Below I list some of my concerns and I hope that the authors find my suggestions helpful.

- • *Language* – (minor) In general the writing is fair, with only minor grammatical errors here and there (e.g., misspellings, improper tenses). However, the cumulative

effect of these minor errors compounds the larger problems the reader faces when trying to understand the methods and results.

We have carefully gone through our revised manuscript to remove these errors

- *Genetic lineages* – The authors don't clarify at the outset why two different populations of birds are tested. As such, the reader is left with the assumption that this was done merely because they were available, rather than to address a specific question. Perhaps move some of the points raised in the discussion regarding this factor to the introduction.

Thanks for pointing out that this creates some confusing to the reader. We provided a brief explanation for our choice of the two selection lines in the discussion. In short, we choose these lines for their variation in foraging behavior and the potentially differential impacts of noise. Furthermore, some of the behavioral differences between these lines have a genetic basis, allowing us to assess whether there is potential for evolutionary adaptation to noise. We hope our choice for the two selection lines is clear now. We have added some lines to the introduction and provide two more references in response to your comment.

- *Methods*

[l., 152] The reason behind using black and white colors is never made clear.

Birds were trained only on conspicuous prey targets during the first phase and we wanted to avoid that this experience would influence our test results. We therefore switched to another conspicuous color. Additionally, we wanted to control for potential effect of ink, which is why we use both plain white paper and black-printed paper for our conspicuous training and test stimuli. We added a line to the method section to explain this.

[l., 154] The point at which training ends and the experiments begin seems arbitrary. Be precise.

Individual birds varied greatly in how they explored the training and test room, how they approached the board, how long they took to land on the board or perch etc. We therefore had to tweak the training to each bird individually, making it hard to come up with a general criterion. We did define *a priori* that they had to grab the prey at least *one time* during training. Furthermore, testing was always carried out one or more days after training, to control at least a little bit for these different training trajectories. We now specify this in the methods.

[l., 185] define pink noise for the reader

We now do.

[l., 205] I found the application of the noise treatment here a bit confusing, given that you previously stated (l., 197) that subjects were acclimatized to the noise treatment while still in their cages. How then, does each trial start with 30s of noise playback that gradually increases?

The cages were adjacent to the test room so non-focal birds were always exposed to the noise during trials with focal birds. We therefore started an experimental day with 1 hour of noise to let them all acclimatize. We think it is clear from the methods.

[l., 208] Explain why you defined an ‘attack’ the way you did. Presumably, the birds usually did not try to eat the target but merely inspected it.

We have phrased this differently now. Most birds would physically attack the target and eat the reward. However, some would land on the nearest perch, some would hover in front of it.

Background manipulation – In general, the presentation of two different tree images with different backgrounds way an important source of confusion for me in these experiments. Line 215 states that 2 printed trees are used during the experiments, so presumably you mean all three experiments.

Correct, we now make sure this is clear in the methods

Why present the birds with two trees and why must they have different backgrounds?

We always presented them two trees to search for targets in order to make sure they were paying attention to the target and not simply flying to the board and hop from small perch to perch until they hit the prey. We started doing this during the training phase, where the targets were presented on only one of the two trees. During the test phase we presented targets on both trees for experiment 1, to avoid any influence of a side bias. In experiment 2 we first put the targets on both trees and repeated the experimental trials with a target placed only on one tree. This last round of trials allowed us to assess whether birds really went for the target and not for the background. In experiment 3 we wanted to give them a choice between targets and thus also used two background trees. The fact that they chose the most conspicuous targets most often in this experiment was again another confirmation that they were trained to search for the target and not something else. We have added some lines to explain this in the methods.

We used different backgrounds as an additional way of avoiding pseudo-replication (not testing birds on targets from just a single tree on a background from just a single tree). Although this introduces additional ‘noise’ in the data, we argue that this makes our results more robust (independent of the unique patterning of a single tree). We have added a line to explain this.

Given that crypsis is clearly a function of the color and patterning of an object relative to its background, the background is clearly an important component. Yet, the reader never gets a sense for how the authors are varying the backgrounds, aside from stating that you presented 2 of 8 possible images during the experiments. Are you randomly selecting two during each trial or are you just using the same two across all trials (the language suggests the latter)?

We have clarified the use and randomization of the background trees in the methods.

Importantly, there’s no discussion of how changes in the background influence the response metric. For example, when you use colored or cryptic models (experiments 2 and 3) are these presented equally on both types of tree images? Do you randomize the positions of these background images or are they always on the same side? It is not apparent that you controlled for this added source of noise in your analyses. Even if you match the color target and the cryptic target to their particular tree within a trial, are you balancing the matching of these targets on each tree image? Given that tree background is neither a fixed nor a random effect in your models, it’s not clear to me how you account for this factor. How do you know that the background itself had no effect on their behavior?

First of all, all targets were tested on the same set of two trees on an experimental day. We randomly replaced background trees between days and not between trials in order because of time constraints. Each individual was thus tested on the same background, so the specific background tree could not explain the observed differences between individuals or between trials on a given day. Importantly, we did not test targets on the background they were created from (cutting triangles out of the tree image in Photoshop). In experiment 2 we tested birds twice on the same combination of stimuli (3 targets x 2 noise treatments). In the first round they were tested on two of the same targets, placed on either one of the trees, their column and row numbers matching. In the second round (on a different experimental day) they were tested on just one tree, the side changed randomly (but not balanced) between trials. This latter round reinsured us that birds were actually paying attention to the target and not just flying to one of the trees to get a closer look. We have made this clearer in the method section.

If two trees are always being presented, why not provide an analysis on individual choice in each experiment, rather than just in the last experiment?

Only in the last experiment the birds were forced to choose between two different targets, placed on either side, again their row and column numbers balanced to avoid any confounding effects of placement. In experiment 1 and the first round of experiment 2 we wanted to avoid overruling effects of a particular side bias (something that often causes problems in behavioral experiments). We therefore presented targets on both trees. Preference for one background or the other could therefore not explain the results during these trials. In the second round of experiment 2 we only presented the target on one side. Birds that made a clear attack on the target always approach the tree containing the target first, so again any preference for 1 tree, or side cannot explain results from this round.

At a minimum, a binomial test would indicate whether the birds explored the tree image containing the target more often than chance could explain. In addition, the presence of two trees raises the question as to what data are used to calculate latencies. Are you only analyzing data from trials in which the birds selected the tree with the target? This seems obvious for attack latencies, but what about approach latencies? Please be more specific.

See our reply above. We now specify this in the methods and results.

o [l.227] I'd consider 17 subjects a modest, not large, number of subjects. Fair enough, we removed large.

Also, what are 'exploration scores'? Do you mean your approach and attack latencies you describe later? This should have been selection line. We have changed this.

o [l. 236] Why is this particular bird's behavior problematic? If you randomize the location of the target each time (l., 221), shouldn't that control for how long they take to 'attack' a prey item once they've approached the board? Unless data from this subject comes out as an influential outlier during your model validation phase, I don't see why you'd discard this bird's data. This particular behavior was problematic because the bird could assess the target from the side, thus not viewing the cryptic pattern. As a consequence, its responses hardly varied across trials. Given that it did not make use of the visual cues we were interested in during its attack we thought it'd be best to exclude this bird from the analyses.

However, the data from this bird did not influence the overall conclusions, so we decided to include it and have updated the stats and figure 2 accordingly.

o [l. 241] – The end of this sentence is confusing '...but balanced across individuals'. Couldn't you just state that you ran a balanced full factorial design? Yes, thank you!

o [l. 246] In general, I feel that you can do a better job of presenting the rational and methods for this last experiment. In the introduction you state that the reason for providing subjects with choices of varying difficulty was to 'ensure that birds relied on visual information during their attack and approach behavior' (l., 116). Yet, the individual and interactive effects of auditory noise and visual information seem to be components of experiment two, not three. Upon reading the description of experiment three on line 246, and your interpretation of these results in the discussion,

"it seems that this experiment is more about how auditory noise can influence choice between targets whose relative visibility varies" exactly - a harder task and one worth pursuing, but not clearly explained. Thanks for the suggestion, we have tried to clarify the rational behind this experiment a bit better.

• *Statistical analyses* - I have some concerns with the validity of the tests conducted that could use some clarification by the authors. Basically, you are fitting complicated hierarchical models to small data sets and this raises some concerns over the validity of your significance tests. I admit that I have much to learn about hierarchical modeling, but I do feel that my experience with mixed-models is fairly representative of ethologists/ecologists in general.

We appreciated and share your concern of fitting complex hierarchical models to data obtained from a moderate number of individuals, which is why we spent a considerable amount of effort exploring different statistical approaches. One approach is to construct a set of candidate null models that vary in their random structure. Comparing these models using an information theoretic approach (in our case compare AIC-values corrected for small sample sizes), as well as assessing residual variation for model assumptions allows one to select the best null model (excluding fixed effects of interest, in our case target and noise treatment). We have added information on the models that we have fitted in this exploration phase to the method section. Furthermore, we now report the outcome of this selection procedure in a supplementary document.

o You state that "various random structures" were fitted [l. 270], but you only provide one. You have also chosen an interesting way to account for individual variation by using perch as your random intercept and subject id as the random slope. I'm not sure that this is the wrong way to model your data, but it is a bit odd and unconventional, so it merits more explanation than you have provided.

We now provide our reasoning and a citation to explain our model selection procedure. Furthermore, we provide the output of this procedure as supplementary information. When comparing the simplest model (birdID as random intercept), with the model that best fitted our data (based on their relative AIC values) we clearly see that a model with latency regressed over order of presentation per individual has the best fit. This can be explained by the observation that birds varied greatly in their latencies across trials on a given day, but each individual varied in its own way. Some were very fast on their first trials and seemed to lose motivation towards the end of the day. Some showed an opposite pattern. This variation that appeared to be independent of our two treatments (which were randomized across order) is captured by the random slope model.

When comparing the output of the simple and best random structure model we also noted that for the simple model all fixed effects were highly significant, whereas for the best model

only noise was still significant. We are therefore more confident that with the latter model we minimize type II errors.

Finally, in response to your comment and reading up on some of the limitations of random effects we have realized that modelling perch row/column as random effect didn't make sense and we reran our models without it. Although *perchrow* explained little variation, removing it caused the effect of selection line to become non-significant. We have adjusted the stats and figures in our manuscript accordingly.

For example, while random slope models are uncommon in ecology, I agree that there is a logic to including both a random effect for the intercept and slope here. However, I fear that you are overparameterizing your models given your sample size (which is modest to begin with and drops over the course of the experiments). Random slope models are known to require a large number of data points to avoid spurious results (See Harrison et al. 2018. PeerJ., doi: 10.7717/peerj.4794 for a recent review). In experiment 2 you have just 9 birds, yet you have a model with two categorical predictors (with 2 and 3 levels, respectively), an interaction between these factors, and two random effects. This model strikes me as running the risk of being overparameterized and providing spurious results. I strongly suggest that the authors provide some supplemental material where the reader can access details supporting the validity of their approach without being distracting from the flow of the paper. Alternatively, you could simply use bird id as a random intercept, which would somewhat simplify the models and account for inherent differences in average performance and issues of independence.

We have now more information on our statistical approach and provide information on simplified models in a supplementary document. Simplified models showed the same result, but seemed to fit the data less well than our random slope model approach, which is why we stick to this model in the main text.

If you wish to retain a random slope, why is perch a random effect as oppose to selection line? I can see how initial perch position could impact a random search, but since you randomized the position of the target across trials then you have already controlled for this nuisance. It seems more biologically intuitive to nest individual id within selection line rather than within initial perch position during approach (Fig. C seems to reinforce nesting id within lineage). Given that your subjects are drawn from two different populations selected for exploratory/bold behavior, one could reasonably expect *a priori* to see differences in latency.

Within these populations, we could then also expect effect strength to vary by individual.

We had not considered modeling individual nested in selection line, which we do in our revised manuscript. However, models with such structure had the highest AIC values (see supplementary info). In response to your comment on perchrow we have decided to remove this factor from the models.

o [l. 280] – If you are still using birds from two selection lines, and your results from experiment 1 show that selection line influences approach latency, why is this term dropped from your model here? I don't see how showing that both selection lines are influenced by noise tells you anything about how they impact individual performance in the visual task.

In our new analyses selection line appears to be non-significant, so we do not need to take this into account for follow-up experiments.

o [l. 281] While I appreciate the intent of experiment three, I found the description of your analysis here rather confusing. Why is noise the only predictor of latency here, but both

noise and target combinations are used to predict choice? How is choice scored here anyway? If you're assuming binomial errors, doesn't that mean that your response is either a 0 or 1? If so, I'm a bit confused by how you score the results from a given trial in which two different targets are presented. On line 275 you state that you scored choice *vs.* no choice. Since you only present an analysis of choice in your third experiment, I'm not sure what this means.

We have redone the analysis by focusing on attacks (y or n) on a predetermined target (the one on the right tree) and by modelling whether the probability of attack depended on the target' pattern as well as the noise treatment. We have rewritten this section on the effect of crypis and noise on attack preferences in the methods and results and hope it is clear now.

A (perhaps) minor point here that relates to the above. Line 276 you refer to using the cbind function to compare choice. Do you mean the function in R to combine values of vectors, columns, or matrices? If so, I'm not sure how your use of this tool clarifies how you scored your response variable for experiment three.

o [l. 284] The number of birds keeps changing between experiments. Please provide an exact count for each experiment when you introduce them, rather than making the reader go back and calculate how many birds we are down to now or wait for the statistics.

We do so now.

• Results

o Fig 1c – why show only the approach latency data for this combination of fixed factors? Even if you found no effect of selection line on attack latencies, the reader is left wondering what those responses looked like. Also, the figure caption says that the boxplots are of the predicted values (model estimates). Is that right, or do you mean the actual data?

Approach and attack latency correlated strongly throughout our experiments, which is why we did not see the value to plot this data for each experiment. We do show approach for Fig. 1 and attack for Fig. 2 and are confident this should provide enough info to interpret our results. Alternatively, we could also include additional figures, but we are already at the page limit.

o [l. 304] 'comparable numbers from both selection lines' is too vague. Be precise.

We added the numbers

o [l. 304-308] I found this confusing. Are days 2 & 3 in figure 1D showing results during training runs? Why?

No, they show results from the test days, we added some more information to make this clear.

o Figure 2 – I didn't find this figure particularly informative. The text is primarily reporting average differences in approach and attack latencies among levels of crypsis. This information is lost in figure 2, which instead focuses on showing individual-level variation

rather than group level differences. Also, why do you again not show results for both approach and attack latency data (particularly since both are apparently influential here)?

Our main aim with figure 2 is to show that there is an interaction between noise and visual information processing. We feel that by showing the individual data we best get this across. Adding boxplots would not add much extra and combining data in a single panel would make it more difficult to focus on the interaction.

o Figure 3 is referenced when presenting the results of the third experiment, but this figure did little to help reduce my confusion. For example, [l., 325] 'a post-hoc analysis revealed that birds preferred to attack the conspicuous prey over the cryptic prey (white *vs.* cryptic, ... Fig 3). White *vs.* cryptic is a level in your fixed effect, so if this combination influenced choice then what was the choice (white or cryptic)? Figure 3 shows the proportion of times that each level of crypsis was selected but doesn't convey how these number vary due to the choice presented. That is, you are not merely asking what the probability of selecting a white target is, but what is the probability of selecting a white target when a colored one is present or when a cryptic one is present? If there is a nearly 60% chance that a bird attacked a white target, how is that number affected by whether the white target was presented with a colored one or the cryptic one? Again, the methods and analysis of this experiment need to be clarified.

We realize that the analysis of this experiment was confusing so we have now used a different approach. The choice of target depends on the target itself as well as the non-chosen target (color vs white is different than color vs cryptic), which creates a level of complexity. We now make this more explicit in the way we modelled the choice. Additionally, we added noise as another factor that could influence this choice. The associated method and result sections have changed accordingly.

o [l. 335- 340, Fig. 1D] – While it is worth pointing out to the reader that auditory noise consistently impacted both approach and attack latencies across these experiments, I don't see how you can pool the data from these different experiments and tack on a mysterious P-value. This seems ill-advised and devalues what is otherwise a perfectly legitimate observation.

We agree that there is no point in providing some sort of test statistics on something that is already clear from the other results. We did aim at drawing attention to the fact that noise repeatedly impacted foraging latencies, but we now remove the test statistics.

• Discussion
o [l., 351] 'noise reduced foraging for conspicuous prey on the first trial, but not subsequent

trials'. It's unclear what you are referring to here.

We refer to the difference between experiment 1 and 2, which we now clarify in the text.

o [l., 382] – This statement raises an interesting point and you seem to offer an explanation based on this concept of habituation. However, in going back through your results I don't find any mention of any analysis to support this statement *vs.* getting these results by chance.

We agree that clear evidence of habituation to simple stimuli would require exposing birds exactly the same number of times to different stimuli to test for differences in rate of habituation. We therefore phrase this carefully by stating 'at least for simple tasks' and by using 'suggest'.

o [l. 387] Given my confusion regarding your analysis of experiment three, this statement about higher-level cognitive effects is not convincing.

We have provided more information on experiment 3 and toned done the argumentation for this statement.

Referee: 2

Comments to the Author(s)
How anthropogenic noise affects wildlife is a research topic to have received considerable attention in the last 10-15 years. One issue that has been particularly well-explored is uni-modal interference – i.e. how noise masks acoustic cues/signals or distracts receivers from making appropriate responses to them. By contrast, there has been relatively little testing of cross-modal or cross-sensory effects – i.e. how noise might interfere with the processing of information from, for example, visual or olfactory cues/signals. This paper reports on a series of simple, but elegant, aviary-based experiments with great tits, investigating the effect of additional noise on foraging for prey of varying levels of conspicuousness. The experiments are well-designed and controlled, the analyses appropriate, the findings are novel and of broad appeal, and the paper generally well-written; I enjoyed reading it. I have just some relatively small points which I help will improve still further this interesting work.

Thanks for your kind words

Lines 31-43: This is ambiguous. It could imply that the birds were trained during the presence or absence of noise; instead, I think the birds were trained in ambient conditions and then tested in either the presence or absence of noise. So, a little rephrasing here would help.

Thanks for catching, we have rephrased it.

Line 51: Whilst it is certainly true that initial work in the field of anthropogenic noise impacts on wildlife had a definite focus on masking of acoustic cues and signals, a large number of studies have now explored a range of other behavioural, physiological, developmental and fitness effects arising from the impact of noise as a distractor and/or stressor (see Shannon et al. 2015 Biol. Rev. for what was a comprehensive list of relevant papers at the time).
True, but it is still generally assumed that these physiological and behavioural responses arise through an impact on acoustic pathways. Which is why we contrast it here with impacts on visual pathways.

Lines 53-54: I would also include mammals (marine and terrestrial) in this list. For example, Parks et al. 2010 Biol. Lett provides an example of a change in whale vocalisations in noise, whereas Kern & Radford 2016 Env. Poll. is an example of how terrestrial mammal communication can be negatively affected by noise.

We have added a reference to mammal work.

Lines 61-64: First sentence puts the focus on ‘non-acoustic’ impacts of anthropogenic noise, but second sentence talks about masking within the acoustic channel. Unless I am missing something, these therefore feel at odds with one another.

The second sentence talks about multimodal communication, which is in part non-acoustically. We can however understand the confusion, and changed non-acoustic to cross-modal.

Lines 84-86: Luo et al. 2015 Glob. Change Biol. and Zhou et al. 2019 Fun. Ecol. provide two examples where there has been a specific consideration of potential mechanisms by which noise can have an impact and so might be worth mentioning as exceptions here.

Are these also on birds? Not on mammals and frogs?

Lines 101-102: Mention of these two selection lines relating to exploratory behaviour comes rather out of the blue and, indeed, this aspect of the paper is the one that is not really set-up or discussed properly. I would suggest that if it were to remain as part of the results, a little introduction to intraspecific variation in responses to noise in general (see Harding et al. 2019 Behav. Ecol. for a recent review on this topic), and exploratory behaviour / personality specifically (see Naguib et al. 2013 Anim. Behav. for a personality-related effect of noise), would be useful (as well as a suitable prediction about what might be expected with respect to this difference).

We have added some more explanation and added two references, see also our reply to reviewer 1 who brought up the same issue.

Lines 110-117: It is not completely clear here how the second two experiments differ from one another; a little extra clarity would be useful.

We have added information to the methods and result section

Line 131: What was the split in terms of sexes and selection lines among the 17 birds that were successfully trained?

We have added information on the numbers

Lines 154-155: Was there a set / formal criteria here (e.g. 10 approaches out of 12)?

Due to the large behavioural differences in approaching and attacking the targets we simply settled for birds that attacked once during the long training day. We have explained the training a bit more in the methods.

Lines 177-180: How many exemplars of each type were created? Different ones for every trial?
Yes, but coming from 1 of 8 tree, which we now specify

Lines 186-188: How many different noise files were created and used?

Only one, as it was artificial and looped continuously.

Lines 204-205: Slight ambiguity as it sounds as if noise levels were gradually increased in control trials, which seems unlikely as those entailed the playback of silence.

Thanks for catching

Lines 210-214: How many trials were abandoned for each of these reasons?

Lines 231-232: Why was this additional training round introduced?

To ensure birds were familiarized with the color and cryptic prey targets as well, which we now explain.

Line 235: What was the sex and selection line split of these remaining 12 birds?

We provide numbers now

Lines 243-244: It is not completely clear what this experiment entailed, in terms of what was presented during a particular trial and what was presented on day one vs day two. Also, why the change to the presentation of one target instead of two on day two?

We have clarified this

Lines 253-255: Why was there scoring by the experimenter if there was a blind scoring option available; the latter is surely the better bet?

We did both, as sometimes scoring on video was not possible and in a few cases the video program crashed during recording. When the results were ambiguous we went for the scores from the blind observer.

Line 264: What % matched (more useful than saying “mostly”)?

Less than 2%

Line 265: How many cases? Fewer than 5%

Lines 278-280: Given there is mention of the number of males and females earlier in the Methods, why was sex not included as a potential fixed factor?

We were not *a priori* interested in sex difference and already reached the limits of our models.

Lines 352-354: These sentences confused me on first reading because the final section of the Results is the only part to mention habituation explicitly… and there the emphasis is on a lack of habituation, which is also the message from Figure 1D. So, a bit of extra clarity here would be useful.

Line 360: It feels like it would be useful to include a few sentences on how the current findings relate to previous work examining foraging and cross-modal/sensory impacts of noise on wildlife; the Discussion is currently a little light on information relating to the main results and relevant literature.

That is partly because there is little work on it, mostly by one of the authors. In view of the page limit we rather restrict the discussion to the main finidings.

Lines 390-393: As with the Introduction, if the comparison of selection lines is to be retained then it needs to be given a little more airtime in the paper – here, that could be how the finding relates to previous work on intraspecific variation in noise responses.

We have added a few lines about the rationale behind the lines, but in light of the page limit don’t want to extent the discussion more than necessary.

Lines 403-406: There are now several studies that have shown differences in responses of animals to noise dependent on previous experience/exposure, which are directly relevant here (see Harding et al. 2019 Behav. Ecol. for table of relevant examples).

Thanks for pointing us towards this interesting paper! We have included it.

Figure 2: Given the first sentence of the figure caption compares across the three panels, it might be useful if those included mean values for ease of reader comparison.

We now report this in text and supplementary information.

Figure 3: It might be worth considering including both noise and quiet values for each of the three moth types, thus showcasing the lack of sound-treatment effect on the probability to attack.

In response to your and reviewer 1’s comments we redid the analyses.

Appendix C

Dear editor(s),

We thank the reviewer and the editor for their effort and appreciate the opportunity to resubmit our manuscript. We have given experiment 3 careful rethinking after reading the comments from reviewer 1. We agree that we did not properly test one of our hypotheses related to this experiment. After rerunning some of the models based on his/her advice we also concluded that we did not have sufficient power to test the predictions of this last hypothesis. We have changed the methods, results and the discussion accordingly. We do however like to point out that all our conclusions were based on the results from experiment 1 and 2, so none of these changes did not alter the conclusions of the manuscript.

Below we provide a detailed answer to the issues raised by the reviewer.

We hope our manuscript is now suitable for publication in Proceedings.

Sincerely,

Wouter Halfwerk & Kees van Oers

Comments to the Author(s).

In general, I have enjoyed reading this study and I found that the authors addressed most of the points brought up by both reviewers and I have only a few minor lingering replies. However, my concerns over the analysis and interpretation of experiment 3 remain. The third experiment is effectively about testing for evidence of cross-modal sensory disruption that impacts a higher-order cognitive process, making a decision – an intention that certainly increases the impact of the study. Based on my interpretation of the descriptions provided, I do not agree with some of the assumptions and so I find myself questioning the results. The description of the new methodology for this experiment remains unnecessarily confusing, as details of the statistical approach vary between sections of the manuscript. I believe that the authors should re-evaluate their approach regarding experiment three and provide some biological arguments for how they choose to proceed. Below I provide details to explain my position and hopefully provide useful feedback.

As a result of your questions and suggestions we became aware that the experimental design of experiment 3 was actually much more complex than anticipated beforehand. Experiment 3 served two purposes. First, it allowed us to test whether individual birds had actually paid attention to the visual patterning, or whether they simply flew to the board and start to hop from perch to perch. We have added a line to the intro and methods to explain our rational. Second, we thought that it allowed us to test what the effect of noise would be when birds had a choice.

When carefully rethinking what we actually tested we now realize that our design cannot discriminate between the effect of noise on detection and the effect of noise on a choice that follows after detection. For example, when birds were presented with the white and cryptic target (creating the largest stimulus contrast) all would attack the white target. See our new figure 3. We therefore do not know whether these birds actually had seen the cryptic target, and so whether they were forced to make a choice or not. The same issue arises when assessing the latency data for experiment 3. Generally, animals take longer to respond to stimuli when they are given a choice, however we do not know whether birds were actually choosing to attack the most conspicuous target, or whether they failed to detect the more cryptic one. Interpreting the latency data thus

turns out to be even more challenging than anticipated. We added some lines to the discussion on experiment 3 to highlight these issues.

We agree that an impact of cross-modal interference on higher-level processing would be very interesting. In our discussion we suggested that noise could influence decision-making, something we agree is an overstatement when based on the results of experiment 3. We have rephrased this section of the discussion.

Minor comments

AIC - You state that you use AIC corrected for small values. Do you mean AICc? I ask because you report AIC everywhere.

Thanks for catching, we have changed the values in the supplementary information accordingly.

Overfitting - AIC assesses relative fit among competing models, but it does not address the question of overall fit. In fact, AIC is biased towards selecting overfit models. You cite Zuur et al. 2009 as one of your references. I think you'll find that Zuur generally encourages simulating ones' final model to assess fit. If you have an overfit model, your simulations will show it because the (original) overfit model was mostly fitting the noise rather than the general trends in your data. The glmmTMB package has made the simulation of glmm's pretty straightforward.

Thanks for sharing your concern regarding overfitting and pointing us to an interesting package. Our raw data shows quite a good match with the model fit (e.g. Fig 1 C), so we are not too concerned. Additionally, our simplest models yielded similar results.

Pseudo-replication - On line 220 you state that 8 different oak trees were used as sources for your target stimuli and background prints to 'avoid pseudo-replication'. I don't see the point of this statement since any concerns over pseudo-replication apply to the repeated measurements on the same birds (within subject) not on whether or not they were tested on the same tree background. In any case, you account for this with your random effects.

We refer to pseudo-replication in stimulus presentation, not in terms of repeated measures per individual. We have removed the statement though to avoid confusion.

Major comments: experiment 3

I don't believe you are correctly testing your hypothesis here and so your statistics do not support your conclusions. Discerning this was more effort than it should be. In the methods [l. 307-312] you describe a model to test your hypothesis, while in the SI [SI, Exp. 3] you present a different model, and your wording in the results [l. 364-370] suggests that you actually applied both models separately. If that is the case, I do not agree with this approach for two reasons: i) it ignores any interactive effect between each of your targets on bird choice and ii) your noise treatment seems unbalanced.

We apologize for the inconsistencies in reporting on experiment 3. As a result of your questions and suggestions we became aware that the experimental design of experiment 3 was actually much more complex than anticipated beforehand. See also above.

The hypothesis being tested here [l. 114-116] is that noise will influence a bird's choice when faced with targets differing in their level of visual crypsis. Ignoring the noise treatment for a moment, an

individual's choice is not only dependent on the level of a given target's crypsis, but also on the opposing target's level of crypsis (you state this yourselves in one of your responses). The model you describe in the methods [l. 308-312] suggests that you actually did something like Y1 ~ T1 x NoiseTreatment + T2 + (Rand. Eff) + GlobalError, where Y1 is the attack score on target 1, T1 and T2 are the corresponding characteristics of the targets (white, colour, or cryptic), and noise is the presence/absence of your sound treatment (+ your random effect and global error terms). If you simply add target 2 as another fixed effect (I think the term covariate is incorrect here), then you are assuming that both targets are independently influencing a given bird's decision to attack target 1. This is equivalent to stating that the chances of attacking a colour target is independent of there being either a white one or a cryptic one on the other 'tree' - an assumption that strikes me as incorrect. You need an interaction between targets. All this being said, you may reach the same conclusions if the interaction term is not significant, but the targets themselves each are. The difference would be that you didn't explicitly ignore the interactive potential, but tested for it and found no evidence of it.

Given the above, how you model the noise should be reconsidered carefully and defended. Following precedence from experiments 1 & 2, one may expect you to include an interactive effect of noise just as you did with latency (in fact, this is what you were doing, albeit incorrectly in my opinion). However, this may be problematic and it's not apparent to me whether or not it is the correct approach. It would, however, be the most parsimonious approach since, by your own words (397-399) don't know how noise is disrupting the decision process. Unfortunately, by accounting for the potential interaction between targets you would be creating a 3-way interaction between each target and the noise, which can be challenging to fit and difficult to interpret. If this approach works, however, it would provide some insight to your point on lines 397-399 in the discussion.

We have used several attempts to remodel the choice data from experiment 3 in response to your comment. The 3-way interaction you suggested failed indeed to run. However, simply scoring the actual attack preference, as a function of the non-chosen target turned out to be the best approach. So, choice~nochoice*noisetreatment, where both choice and nochoice would be one of three possible states (white/color/cryptic). An alternative method was to model the attack (0 or 1) of a predetermined side (left or right target) as a function of the level of crypsis of that target. As fixed effect we added experimental group (one of three possible stimulus combinations). This method yielded a similar result.

Your approach with latency in Exp. 3 suffers from the same issues outlined above for the analysis of target choice. However, not having to specify a particular target as your response variable would seem to allow for additional latitude on how you approach this problem. For instance, if a 3-way interaction proves to be too much for your sample size, couldn't you use the stimulus pairing treatment as an index of relative difficulty (supported by your results in Exp. 2)? If so, you might consider simply regressing latency on the interaction between stimulus treatment x noise.

See our former response.