THE ROYAL SOCIETY
PUBLISHING

PROCEEDINGS B

Categorical colour perception occurs in both signalling and non-signalling colour ranges in a songbird

Matthew N. Zipple, Eleanor M. Caves, Patrick A. Green, Susan Peters, Sönke Johnsen and Stephen Nowicki

Article citation details
Proc. R. Soc. B **286**: 20190524.
http://dx.doi.org/10.1098/rspb.2019.0524

Review timeline

Original submission:	2 January 2019
1st revised submission:	4 March 2019
2nd revised submission:	18 April 2019
3rd revised submission:	7 May 2019
Final acceptance:	7 May 2019

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

Review History

RSPB-2019-0009.R0 (Original submission)

Review form: Reviewer 1

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

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

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

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

Is the length of the paper justified?
Yes

Should the paper be seen by a specialist statistical reviewer?
No

Do you have any concerns about statistical analyses in this paper? If so, please specify them explicitly in your report.
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?
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
The manuscript by Zipple and colleagues is a follow-up study of a study published 2018, in which the authors argue for categorical colour perception in zebra finch females discriminating colours in the orange-red range of colours that zebra finch beaks can take. Here, the authors present similar data and make the point that the same zebra finch females also have categorical colour perception in the blue-green range, even though the effect is smaller.

I think the data as such are solid, and the paper is well written. However, as it will be widely cited as proof for a new perceptual capability in birds, I think higher demands should be put on clarification of the data analysis. I am not arguing that data analysis is incorrect in any way, however, method are new and need to be compared to classic methods. The two main points (detailed below) are that (1) colour distances - except in the first diagram - are reduced to colour steps although they are different in size, and (2) the criterion for passing a trial are two correct choices out of two, which ignores single correct choices completely. Third, it would be important to better proof that the UV cone of the birds did not make a contribution and thus really can be ignored in the analysis.

I am sure the authors have all the data to convince me (and other readers) that their methods give the same results as would classic methods, and that their conclusions hold, so I think it would be worth the effort.

1. Colours and colour measurements
Clearly, if a trichromatic colour space is used for a bird, one wants to be sure that UV does not play any role - otherwise, an unexpectedly larger colour difference due to UV reflectance could completely change the situation. This problem is slightly larger in the blue-green range than in the yellow-red range, as yellow and red pigments used for printing always (or almost always) have high pass characteristics with respect to wavelength, while green and blue pigments have band pass characteristics which may easily extend into the ultraviolet.

Thus, it is a bit worrying that one of the light measurement curves is clearly off but still has been used in the average curve. Curve 0010_4 represents about a factor of ten lower light intensity. Was this used? Why is it lower? The remaining curves show that the light source does have some UV component (more importantly, some light visible to the UV cone), even though a low one.

Even more worrying, some curves, at short wavelengths, have negative components - this is a sign that dark noise is not taken out fully. This could be caused by warming of the lamp during the measurements, but it also means that there could have been UV available. Some clarification is needed here.

In addition, it seems that the sensitivity of zebra finch cones for wavelengths below 400 nm have not been taken into account, as they are not included in the curves in the supplement, and are the reflectances at wavelengths below 300 nm are not given either. They would help to convince the reader of the statement that there is no "significant" reflection in that range - what ever significant means here.

In sum, this issue should be dealt with to convince the reader that the colour calculations are indeed correct, and no hidden UV signal is responsible for the difference in discriminability of different colour pairs.

2. Scaling of the model colour space
this is really a comment rather than criticism: making a difference between delta s and JNDs and defining one as 2,3 time the other is really only an academic exercise and irrelevant, as long as the real lever of receptor noise is not measured. You could equally well adjust the assumed noise level in the zebra finch photoreceptors by a factor of 2,3 and use 1 as the threshold. It may be important not to forget what model calculations are based on.

3. Pass criterion
It would be good for a general reader to understand what "pass" really means in terms of statistics. So it would help to give what is the statistical probability that a bird chooses the two positive stimuli first, if she chooses by chance? the probability of the first correct choice is 1/3, and the second choice is 1/5 so in total, the chance to do both right should be 1/15 - correct? How often would that happen in 10 trials? Thus, what pass frequency is statistically different from chance? Why do trials in which the bird chooses one correct but a second incorrect disk have the same value as trials in which a bird chooses two incorrect ones? How would the data look like if, for instance, the % correct choices for each trial was taken for all 10 trials? This would allow a general reader to compare these data to the data obtained in other studies with more classical methods.

A simple case can illustrate my concern: think of a bird that makes one correct and one incorrect choice in each of 10 trials. If a bird makes 2 choices per trial, she would then have 10 correct choices out of 20. In your analysis, this bird would be counted as 0% pass. another bird makes 2 correct choices in 5 out of 10 trials, and no correct choice in the other 5 trials, adding up to the same number of correct choices, 10 out of 20. However, this bird would be counted a 50% passed trials. I am not saying both birds have the same performance, but I would argue that the difference may be more subtle than your analysis lets the reader think.

4. Colour distances and colour steps
Even though there the authors make a correct statement that the best discrimination did not happen for the smaller colour steps, it would be very helpful to plot the percentage of correct choices as a function of the colour difference, not only over each colour step as if they were equal. Other colour discrimination studies often show very steep psychometric functions for discrimination.

5. The contribution of each colour step to pass frequency
This is a rather complicated measure, which, again, does not take into account, the colour differences between stimuli (which, actually, don not sum up linearly if the colour loci are not on a line). Again, it would be helpful to present these data also in a way, in which the specific colour differences are taken into account.

6. Comparison with the previous study
A main difference between both studies was the presence or absence of achromatic differences between the colours. This should clearly be discussed before speculating about and referring to similarities to humans and connections to environments. Basically, the presence or absence of achromatic differences is the most parsimonious explanation that the differences in the strength of the results (if they hold based on a more classic evaluation of results in terms of statistically different choice frequencies in the different tests instead of pass frequencies that ignore single correct choices in trials, as outlined above).

Review form: Reviewer 2 (Misha Vorobyev)

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?
Acceptable

Is the length of the paper justified?
Yes

Should the paper be seen by a specialist statistical reviewer?
No

Do you have any concerns about statistical analyses in this paper? If so, please specify them explicitly in your report.
No

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

Is it accessible?
Yes

Is it clear?
Yes

Is it adequate?
Yes

Do you have any ethical concerns with this paper?
Yes

Comments to the Author
The paper “categorical colour perception occurs in both signalling and non-signalling colours in songbird” is well written and asks an important question. In the previous paper the authors investigated colour categorisation along the red-orange direction in the colour space. This paper investigates colour categorisation along blue-green direction in the colour space.
The authors find that the birds trained to select two coloured disks demonstrate the best performance with the colours 2-3, which is at odds with the prediction based on colour distances and agrees with the hypothesis that there is a categorical boundary between colours 2 and 3. While it is plausible that birds indeed have categorical colour perception, the results can be explained using several other hypotheses. I do not think that alternative explanations can be ruled out. However alternatives need to be discussed.
An alternative to the colour categorisation hypothesis is the hypothesis of proximity judgements. The birds can simply select the colours that are more different (have larger colour distance) from each other. Usually colour distance is calculated on the basis of threshold measurements. The authors estimate colour distance using the receptor noise limited model. This is the simplest model that ignores the later stages of colour processing. Predictions of this model agrees reasonably with the results obtained using stimuli in the vicinity of achromatic point. However I remains uncertain if the model gives correct predictions for stimuli that the authors used. I suggest that the experimental results with birds are discussed in more details.
The distance based on thresholds does not necessarily agrees with perceptual colour distance. Birds also can select the most salient edge. In humans, edge detection is mediated solely by a luminance mechanism. The authors suggest that in birds it is mediated by double cones. However this hypothesis is based on limited experimental evidence. Possibly L and M cones can be involved in the process.
Preference for colour combination also may explain the results.
Minor points
1. The authors state that as ambient light they used illumination A. However experiments were performed under halogen bulbs with similar colour temperature but spiky spectrum. For humans, these two illuminations are almost metameric. However, for birds, quantum catches calculated under these illuminations must differ substantially. Please, calculate quantum catches using correct illumination. The colour distances may change substantially and therefore the conclusions may be affected. It may also be useful to measure the spectrum of light reflected from Munsell chips illuminated by halogen bulbs.
2. Please indicate in the figure 1 the numbers corresponding to each colour. The Munsell name is less important, you can add it the figure legend.

Decision letter (RSPB-2019-0009.R0)

04-Feb-2019

Dear Mr Zipple:

I am writing to inform you that your manuscript RSPB-2019-0009 entitled "Categorical colour perception occurs in both signalling and non-signalling colour ranges in a songbird" has, in its current form, been rejected for publication in Proceedings B.

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

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

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

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

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

Sincerely,

Proceedings B
mailto: proceedingsb@royalsociety.org

Associate Editor
Board Member: 1
Comments to Author:
Two expert reviewers have now seen your manuscript and both are positive, although both also feel that a significant revision is required before the manuscript is ready for publication. Reviewer 1 is concerned about several aspects of the new methods you employ (particularly the use of colour steps rather than colour distances, and the choice criterion you have used for a bird to pass a trial), and wishes to see an extra analysis of your data using more classical methods (possibly included as supplementary material). Moreover this reviewer is concerned about the possibility of the UV cone making a contribution to your results (a possibility which is currently unaccounted for). Reviewer 2 is largely worried that your results might also be explainable by alternative hypotheses, a possibility that needs to be discussed (and hopefully eliminated) in the Discussion.

Reviewer(s)' Comments to Author:

Referee: 1

Comments to the Author(s)
The manuscript by Zipple and colleagues is a follow-up study of a study published 2018, in which the authors argue for categorical colour perception in zebra finch females discriminating

colours in the orange-red range of colours that zebra finch beaks can take. Here, the authors present similar data and make the point that the same zebra finch females also have categorical colour perception in the blue-green range, even though the effect is smaller.

I think the data as such are solid, and the paper is well written. However, as it will be widely cited as proof for a new perceptual capability in birds, I think higher demands should be put on clarification of the data analysis. I am not arguing that data analysis is incorrect in any way, however, method are new and need to be compared to classic methods. The two main points (detailed below) are that (1) colour distances - except in the first diagram - are reduced to colour steps although they are different in size, and (2) the criterion for passing a trial are two correct choices out of two, which ignores single correct choices completely. Third, it would be important to better proof that the UV cone of the birds did not make a contribution and thus really can be ignored in the analysis.

I am sure the authors have all the data to convince me (and other readers) that their methods give the same results as would classic methods, and that their conclusions hold, so I think it would be worth the effort.

1. Colours and colour measurements
Clearly, if a trichromatic colour space is used for a bird, one wants to be sure that UV does not play any role - otherwise, an unexpectedly larger colour difference due to UV reflectance could completely change the situation. This problem is slightly larger in the blue-green range than in the yellow-red range, as yellow and red pigments used for printing always (or almost always) have high pass characteristics with respect to wavelength, while green and blue pigments have band pass characteristics which may easily extend into the ultraviolet.

Thus, it is a bit worrying that one of the light measurement curves is clearly off but still has been used in the average curve. Curve 0010_4 represents about a factor of ten lower light intensity. Was this used? Why is it lower? The remaining curves show that the light source does have some UV component (more importantly, some light visible to the UV cone), even though a low one.

Even more worrying, some curves, at short wavelengths, have negative components - this is a sign that dark noise is not taken out fully. This could be caused by warming of the lamp during the measurements, but it also means that there could have been UV available. Some clarification is needed here.

In addition, it seems that the sensitivity of zebra finch cones for wavelengths below 400 nm have not been taken into account, as they are not included in the curves in the supplement, and are the reflectances at wavelengths below 300 nm are not given either. They would help to convince the reader of the statement that there is no "significant" reflection in that range - what ever significant means here.

In sum, this issue should be dealt with to convince the reader that the colour calculations are indeed correct, and no hidden UV signal is responsible for the difference in discriminability of different colour pairs.

2. Scaling of the model colour space
this is really a comment rather than criticism: making a difference between delta s and JNDs and defining one as 2,3 time the other is really only an academic exercise and irrelevant, as long as the real lever of receptor noise is not measured. You could equally well adjust the assumed noise level in the zebra finch photoreceptors by a factor of 2,3 and use 1 as the threshold. It may be important not to forget what model calculations are based on.

3. Pass criterion
It would be good for a general reader to understand what "pass" really means in terms of statistics. So it would help to give what is the statistical probability that a bird chooses the two positive stimuli first, if she chooses by chance? the probability of the first correct choice is 1/3, and the second choice is 1/5 so in total, the chance to do both right should be 1/15 - correct? How often would that happen in 10 trials? Thus, what pass frequency is statistically different from chance? Why do trials in which the bird chooses one correct but a second incorrect disk have the same value as trials in which a bird chooses two incorrect ones? How would the data look like if, for instance, the % correct choices for each trial was taken for all 10 trials? This would allow a general reader to compare these data to the data obtained in other studies with more classical methods.

A simple case can illustrate my concern: think of a bird that makes one correct and one incorrect choice in each of 10 trials. If a bird makes 2 choices per trial, she would then have 10 correct choices out of 20. In your analysis, this bird would be counted as 0% pass. another bird makes 2 correct choices in 5 out of 10 trials, and no correct choice in the other 5 trials, adding up to the same number of correct choices, 10 out of 20. However, this bird would be counted a 50% passed trials. I am not saying both birds have the same performance, but I would argue that the difference may be more subtle than your analysis lets the reader think.

4. Colour distances and colour steps
Even though there the authors make a correct statement that the best discrimination did not happen for the smaller colour steps, it would be very helpful to plot the percentage of correct choices as a function of the colour difference, not only over each colour step as if they were equal. Other colour discrimination studies often show very steep psychometric functions for discrimination.

5. The contribution of each colour step to pass frequency
This is a rather complicated measure, which, again, does not take into account, the colour differences between stimuli (which, actually, don not sum up linearly if the colour loci are not on a line). Again, it would be helpful to present these data also in a way, in which the specific colour differences are taken into account.

6. Comparison with the previous study
A main difference between both studies was the presence or absence of achromatic differences between the colours. This should clearly be discussed before speculating about and referring to similarities to humans and connections to environments. Basically, the presence or absence of achromatic differences is the most parsimonious explanation that the differences in the strength of the results (if they hold based on a more classic evaluation of results in terms of statistically different choice frequencies in the different tests instead of pass frequencies that ignore single correct choices in trials, as outlined above).

Referee: 2

Comments to the Author(s)
The paper “categorical colour perception occurs in both signalling and non-signalling colours in songbird” is well written and asks an important question. In the previous paper the authors investigated colour categorisation along the red-orange direction in the colour space. This paper investigates colour categorisation along blue-green direction in the colour space.

The authors find that the birds trained to select two coloured disks demonstrate the best performance with the colours 2-3, which is at odds with the prediction based on colour distances and agrees with the hypothesis that there is a categorical boundary between colours 2 and 3. While it is plausible that birds indeed have categorical colour perception, the results can be explained using several other hypotheses. I do not think that alternative explanations can be ruled out. However alternatives need to be discussed.
An alternative to the colour categorisation hypothesis is the hypothesis of proximity judgements. The birds can simply select the colours that are more different (have larger colour distance) from each other. Usually colour distance is calculated on the basis of threshold measurements. The authors estimate colour distance using the receptor noise limited model. This is the simplest model that ignores the later stages of colour processing. Predictions of this model agrees reasonably with the results obtained using stimuli in the vicinity of achromatic point. However I remains uncertain if the model gives correct predictions for stimuli that the authors used. I suggest that the experimental results with birds are discussed in more details.
The distance based on thresholds does not necessarily agrees with perceptual colour distance.
Birds also can select the most salient edge. In humans, edge detection is mediated solely by a luminance mechanism. The authors suggest that in birds it is mediated by double cones. However this hypothesis is based on limited experimental evidence. Possibly L and M cones can be involved in the process.
Preference for colour combination also may explain the results.
Minor points
1. The authors state that as ambient light they used illumination A. However experiments were performed under halogen bulbs with similar colour temperature but spiky spectrum. For humans, these two illuminations are almost metameric. However, for birds, quantum catches calculated under these illuminations must differ substantially. Please, calculate quantum catches using correct illumination. The colour distances may change substantially and therefore the conclusions may be affected. It may also be useful to measure the spectrum of light reflected from Munsell chips illuminated by halogen bulbs.
2. Please indicate in the figure 1 the numbers corresponding to each colour. The Munsell name is less important, you can add it the figure legend.

Author's Response to Decision Letter for (RSPB-2019-0009.R0)

See Appendix A.

RSPB-2019-0524.R0

Review form: Reviewer 1

Recommendation
Accept with minor revision (please list in comments)

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

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

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?
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
Congratulations! The new version of the manuscript the authors have taken comments into account, and clarified a lot of uncertainties present in the original submission. The authors present new measurements, have calculated their pass criterion in an additional way and show that the result is not changed. However, a lot of the added information can only be seen in the reply to the reviewers, or if the reader plots curves from the data in the supplement.

I therefore suggest to add the plots (e.g. reflectances of the colour stimuli, light measurements and % quantum catch of the UV cone) to the supplementary text file. It costs nothing but helps the reader.

In the spreadsheet, please give the units for each measurement. I doubt it is just “photons” for the light, and it should say percent reflectance for the Munsell chips and also how the cone sensitivity curves are normalized. In the supplement file, abbreviations should be explained. I guess most people know that CI is supposed to mean confidence interval, but it should be said at first mention anyway. Maybe in 50 years from now, CI typically is used for something completely different.

I specifically appreciate the additional analysis of choices using a second, more relaxed choice criterion. My original point was not that the authors should change their criterion. The point was rather that that the reader should be convinced that this choice criterion is reliable, and you show it is – great. Introducing new methods is most helpful if the results are still comparable with those of other studies.

The same applies to the “classic” analysis of the results. What it shows is that the results are much more spread than they are in a classical discrimination threshold test. If jnd calculations are in an acceptable range (no severe over- or under-estimation of noise) then a bird should definitely be at 80% correct choices with 10 jnds – if the bird really does its best.

That this does not always happen immediately leads the reader to think about the reasons and very likely the reason is that the experimental design is such that birds to do show their absolute best discrimination ability. That is fine, as it likely is the condition under which they show signa of using categorization instead. However, I think it is worth mentioning this fact. This may become a philosophical discussion but IF categorical colour perception means that categorization overrides physiology (by which mechanism ever) THEN the conditions under which this happens need to be defined.

Please mention what Table S3 shows, in your results section. You only refer to it, but don’t mention the important fact that the colour step model does explain the data better than the model just taking chromatic distance into account. In the end, this is the most solid statistical proof of the categorization so should clearly be written out as it is in the reply to the reviewers. As with other comments to the first version: most of them were thought to ask you for better prove that your interpretation of data is correct. Strong claims require strict and critical tests, and it is great to see that you can give these, so please present them to the readers! The point with reviewers is that they should be the most critical readers, and play the role of advocati diaboli, to make sure the paper convinces other critical readers as well!

In reply to the previous comment on jnd versus ?S: As I said before, this was just a comment, and I don’t request you to do anything about this but my main point was that this is still an academic discussion as the noise levels in zebra finch receptors have never been measured so nobody knows the “real” absolute jnds for the species. Thus we do not know whether the best behavioural thresholds are higher than these absolute thresholds because we do not know the absolute thresholds.

The most simple suggestion to resolve this and the point made before, could be that in the task used in this experiment, birds do not really perform at their best but the only way to find this out would be do give them the chance to perform as good as they can and that would imply a rather boring threshold experiment. For the present paper, there is need to do this.

In short, my suggestion is that
1. the authors add more of the graphs that they present in the reply to the reviewers, in the text supplement, and that they check that they explain all abbreviations and correct units in texts and tables.
2. The authors should mention that discrimination – according to colour distances – is actually worse in this experiment than would be expected in an experiment that has the goal to find the best discrimination with lowest threshold. It is under these conditions, that categories play a role.

Else, this is now a very exciting manuscript, specifically as it now gives more method-critical details which indeed can convince even a critical reader! I also find the reply to the comments made by the second reviewer very convincing.

Review form: Reviewer 2

Recommendation
Reject – article is scientifically unsound

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?
No

Do you have any concerns about statistical analyses in this paper? If so, please specify them explicitly in your report.
No

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

> **Is it accessible?**
> Yes
>
> **Is it clear?**
> No
>
> **Is it adequate?**
> No

Do you have any ethical concerns with this paper?
No

Comments to the Author
I have two major concerns with the revision.

1. In my previous review, I pointed that the authors calculated quantum catches using incorrect illumination. However, the authors did not change all results using correct illumination. They argue that results change only marginally if the wrong illumination is used. As long as the calculations of quantum catches are not reliable all conclusions are not valid.

In the response to reviewers, the authors present the spectrum in logarithmic scale. This scale is not relevant to calculating quantum catches. I am surprised that the spectrum does not contain spikes around 600 nm. These spikes are likely to result in non-marginal changes in the calculated L cone quantum catch.

In the supplementary materials, the units of illumination are not given. Which units have been used to calculate quantum catches? How was spectrometer calibrated? What was the sampling rate and spectral resolution? How were the 1nm interval data generated?

I suggest to check if the light above 700nm contribute to quantum catch of L cone, because the lamp may have high output in the red- near infra-red part of the spectrum.

To convince a reader that there is no mistake in calculations or measurements, more information must be given in the supplementary materials.

I also suggest that the authors reanalyse the data from previous paper using correct illumination.

2. In my previous review, I suggested that the authors discuss alternative explanations. However the authors preferred not to do it. Instead, they wrote a lengthy explanations addressed to reviewer.
All conclusions are based on the deviations of bird performance from that predicted by the model. This must be clearly stated in the paper. I find this line of arguments to be thin and I am not entirely convinced by the authors' arguments.

Multiple reasons that may explain the deviation of model predictions from experimental results. Readers must be convinced that the deviations cannot be explained by errors in calculations or measurements (see comment 1, sufficient information must be given in the Supplement). Also, the alternative explanations must be discussed in the paper, and readers may decide if the arguments in favour of the hypothesis of categorisation are convincing.

Minor points:
Line 101 wrong format of Eq.

Spectral sensitivities of zebra finch. Lind did not measure the sensitivities, leave this reference for modelling and give reference to msp and oil droplet modelling.

Decision letter (RSPB-2019-0524.R0)

01-Apr-2019

Dear Mr Zipple:

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

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

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

When submitting your revision please upload a file under "Response to Referees" in the "File Upload" section. This should document, point by point, how you have responded to the reviewers' and Editors' comments, and the adjustments you have made to the manuscript. We require a copy of the manuscript with revisions made since the previous version marked as 'tracked changes' to be included in the 'response to referees' document.

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

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

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

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

Data accessibility and data citation:
It is a condition of publication that you make available the data and research materials supporting the results in the article. 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,
Proceedings B
mailto: proceedingsb@royalsociety.org

Associate Editor Board Member
Comments to Author:
Both original reviewers have now reassessed your manuscript and their reactions are both positive (reviewer 1, who recommends acceptance with minor revisions) and negative (reviewer 2, who recommends rejection). Reviewer 1 nonetheless has a number of specific criticisms and comments that still need to be addressed, and interestingly these criticisms and comments are not entirely dissimilar to those of the much more critical reviewer 2. Both reviewers feel that many of your previous responses should be incorporated into the manuscript (which you haven't done), and more information (i.e. clarity) needs to be provided for the light measurements you have made, since these underpin the interpretation of your data and several major conclusions of your study. Both reviewers (and particularly reviewer 2) also fear that your results might be explained by alternative hypotheses, and that while this was discussed at length in your previous responses to their criticisms, this discussion failed to make it to the manuscript.

It is my opinion that all of the comments of both reviewers - despite their very different recommendations - can be accommodated satisfactorily, and that once this has been achieved, the manuscript will be in a fine state for publication.

Reviewer(s)' Comments to Author:

Referee: 1

Comments to the Author(s).
Congratulations! The new version of the manuscript the authors have taken comments into account, and clarified a lot of uncertainties present in the original submission. The authors present new measurements, have calculated their pass criterion in an additional way and show

that the result is not changed. However, a lot of the added information can only be seen in the reply to the reviewers, or if the reader plots curves from the data in the supplement.

I therefore suggest to add the plots (e.g. reflectances of the colour stimuli, light measurements and % quantum catch of the UV cone) to the supplementary text file. It costs nothing but helps the reader.

In the spreadsheet, please give the units for each measurement. I doubt it is just “photons” for the light, and it should say percent reflectance for the Munsell chips and also how the cone sensitivity curves are normalized. In the supplement file, abbreviations should be explained. I guess most people know that CI is supposed to mean confidence interval, but it should be said at first mention anyway. Maybe in 50 years from now, CI typically is used for something completely different.

I specifically appreciate the additional analysis of choices using a second, more relaxed choice criterion. My original point was not that the authors should change their criterion. The point was rather that that the reader should be convinced that this choice criterion is reliable, and you show it is – great. Introducing new methods is most helpful if the results are still comparable with those of other studies.

The same applies to the “classic” analysis of the results. What it shows is that the results are much more spread than they are in a classical discrimination threshold test. If jnd calculations are in an acceptable range (no severe over- or under-estimation of noise) then a bird should definitely be at 80% correct choices with 10 jnds – if the bird really does its best.

That this does not always happen immediately leads the reader to think about the reasons and very likely the reason is that the experimental design is such that birds to do show their absolute best discrimination ability. That is fine, as it likely is the condition under which they show signa of using categorization instead. However, I think it is worth mentioning this fact. This may become a philosophical discussion but IF categorical colour perception means that categorization overrides physiology (by which mechanism ever) THEN the conditions under which this happens need to be defined.

Please mention what Table S3 shows, in your results section. You only refer to it, but don’t mention the important fact that the colour step model does explain the data better than the model just taking chromatic distance into account. In the end, this is the most solid statistical proof of the categorization so should clearly be written out as it is in the reply to the reviewers. As with other comments to the first version: most of them were thought to ask you for better prove that your interpretation of data is correct. Strong claims require strict and critical tests, and it is great to see that you can give these, so please present them to the readers! The point with reviewers is that they should be the most critical readers, and play the role of advocati diaboli, to make sure the paper convinces other critical readers as well!

In reply to the previous comment on jnd versus □S: As I said before, this was just a comment, and I don’t request you to do anything about this but my main point was that this is still an academic discussion as the noise levels in zebra finch receptors have never been measured so nobody knows the “real” absolute jnds for the species. Thus we do not know whether the best behavioural thresholds are higher than these absolute thresholds because we do not know the absolute thresholds.

The most simple suggestion to resolve this and the point made before, could be that in the task used in this experiment, birds do not really perform at their best but the only way to find this out would be do give them the chance to perform as good as they can and that would imply a rather boring threshold experiment. For the present paper, there is need to do this.

In short, my suggestion is that
1. the authors add more of the graphs that they present in the reply to the reviewers, in the text supplement, and that they check that they explain all abbreviations and correct units in texts and tables.
2. The authors should mention that discrimination – according to colour distances – is actually worse in this experiment than would be expected in an experiment that has the goal to find the best discrimination with lowest threshold. It is under these conditions, that categories play a role.

Else, this is now a very exciting manuscript, specifically as it now gives more method-critical details which indeed can convince even a critical reader! I also find the reply to the comments made by the second reviewer very convincing.

Referee: 2

Comments to the Author(s).
I have two major concerns with the revision.

1. In my previous review, I pointed that the authors calculated quantum catches using incorrect illumination. However, the authors did not change all results using correct illumination. They argue that results change only marginally if the wrong illumination is used. As long as the calculations of quantum catches are not reliable all conclusions are not valid.

In the response to reviewers, the authors present the spectrum in logarithmic scale. This scale is not relevant to calculating quantum catches. I am surprised that the spectrum does not contain spikes around 600 nm. These spikes are likely to result in non-marginal changes in the calculated L cone quantum catch.

In the supplementary materials, the units of illumination are not given. Which units have been used to calculate quantum catches? How was spectrometer calibrated? What was the sampling rate and spectral resolution? How were the 1nm interval data generated?

I suggest to check if the light above 700nm contribute to quantum catch of L cone, because the lamp may have high output in the red- near infra-red part of the spectrum.

To convince a reader that there is no mistake in calculations or measurements, more information must be given in the supplementary materials.

I also suggest that the authors reanalyse the data from previous paper using correct illumination.

2. In my previous review, I suggested that the authors discuss alternative explanations. However the authors preferred not to do it. Instead, they wrote a lengthy explanations addressed to reviewer.
All conclusions are based on the deviations of bird performance from that predicted by the model. This must be clearly stated in the paper. I find this line of arguments to be thin and I am not entirely convinced by the authors' arguments.

Multiple reasons that may explain the deviation of model predictions from experimental results. Readers must be convinced that the deviations cannot be explained by errors in calculations or measurements (see comment 1, sufficient information must be given in the Supplement). Also, the alternative explanations must be discussed in the paper, and readers may decide if the arguments in favour of the hypothesis of categorisation are convincing.

Minor points:
Line 101 wrong format of Eq.

Spectral sensitivities of zebra finch. Lind did not measure the sensitivities, leave this reference for modelling and give reference to msp and oil droplet modelling.

Author's Response to Decision Letter for (RSPB-2019-0524.R0)

See Appendix B.

Decision letter (RSPB-2019-0524.R1)

03-May-2019

Dear Mr Zipple

I am pleased to inform you that your manuscript RSPB-2019-0524.R1 entitled "Categorical colour perception occurs in both signalling and non-signalling colour ranges in a songbird" has been accepted for publication in Proceedings B.

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

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

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

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

1) A text file of the manuscript (doc, txt, rtf or tex), including the references, tables (including captions) and figure captions. Please remove any tracked changes from the text before submission. PDF files are not an accepted format for the "Main Document".

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

3) Electronic supplementary material: this should be contained in a separate file and where possible, all ESM should be combined into a single file. All supplementary materials accompanying an accepted article will be treated as in their final form. They will be published alongside the paper on the journal website and posted on the online figshare repository. Files on figshare will be made available approximately one week before the accompanying article so that the supplementary material can be attributed a unique DOI.

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

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

5) Data accessibility section and data citation
It is a condition of publication that data supporting your paper are made available either in the electronic supplementary material or through an appropriate repository.

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

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

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

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

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

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

Sincerely,

Proceedings B
mailto:proceedingsb@royalsociety.org

Associate Editor:
Comments to Author:
I am now satisfied that you have made all needed changes to the manuscript to satisfy the great majority of comments and criticisms raised in the previous revision by both reviewers. I have carefully considered the first major criticism of Reviewer 2 by reassessing your initial response to this criticism from the first version of the manuscript and your latest response to this criticism here. I feel that your view on the matter is justified, although I also think you should include the colour step table from your initial set of responses to reviewers (from the first revision) as a table in the Supplementary Materials (with an extra sentence to motivate this table in the main current manuscript, somewhere around lines 113-117). Regarding Reviewer 2's question about the spectrometer calibration, I also think it would be worth adding to the Methods section the information you gave in your response. Following these adjustments my feeling is that your manuscript will become a valuable addition to the literature.

Decision letter (RSPB-2019-0524.R2)

07-May-2019

Dear Mr Zipple

I am pleased to inform you that your manuscript entitled "Categorical colour perception occurs in both signalling and non-signalling colour ranges in a songbird" has been accepted for publication in Proceedings B.

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

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

If you have any queries regarding the production of your final article or the publication date please contact procb_proofs@royalsociety.org

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.

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

Sincerely,

Proceedings B
mailto: proceedingsb@royalsociety.org

Appendix A

1. Colours and colour measurements
Clearly, if a trichromatic colour space is used for a bird, one wants to be sure that UV does not play any role - otherwise, an unexpectedly larger colour difference due to UV reflectance could completely change the situation. This problem is slightly larger in the blue-green range than in the yellow-red range, as yellow and red pigments used for printing always (or almost always) have high pass characteristics with respect to wavelength, while green and blue pigments have band pass characteristics which may easily extend into the ultraviolet.

Thus, it is a bit worrying that one of the light measurement curves is clearly off but still has been used in the average curve. Curve 0010_4 represents about a factor of ten lower light intensity. Was this used? Why is it lower? The remaining curves show that the light source does have some UV component (more importantly, some light visible to the UV cone), even though a low one.

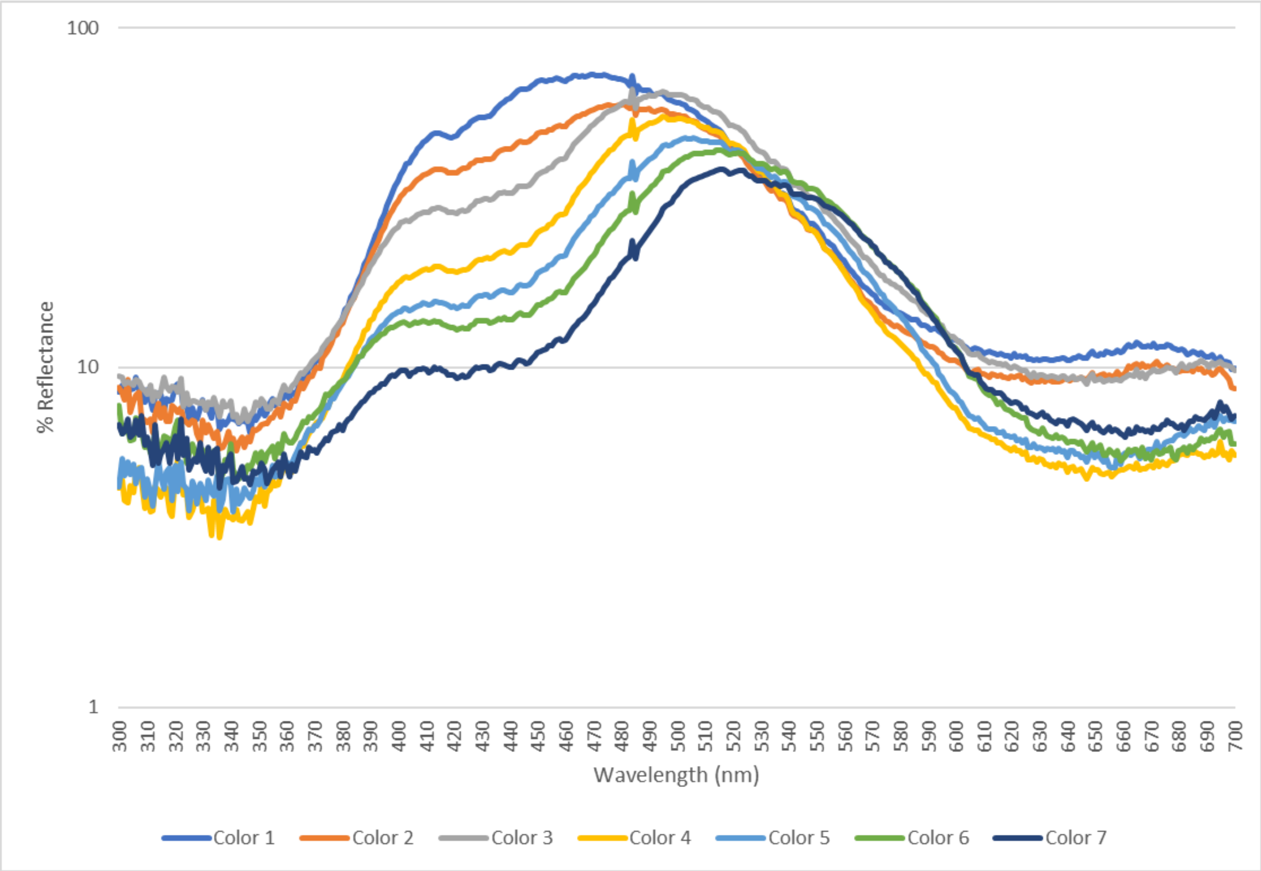
Even more worrying, some curves, at short wavelengths, have negative components - this is a sign that dark noise is not taken out fully. This could be caused by warming of the lamp during the measurements, but it also means that there could have been UV available. Some clarification is needed here.

In addition, it seems that the sensitivity of zebra finch cones for wavelengths below 400 nm have not been taken into account, as they are not included in the curves in the supplement, and are the reflectances at wavelengths below 300 nm are not given either. They would help to convince the reader of the statement that there is no "significant" reflection in that range - what ever significant means here.

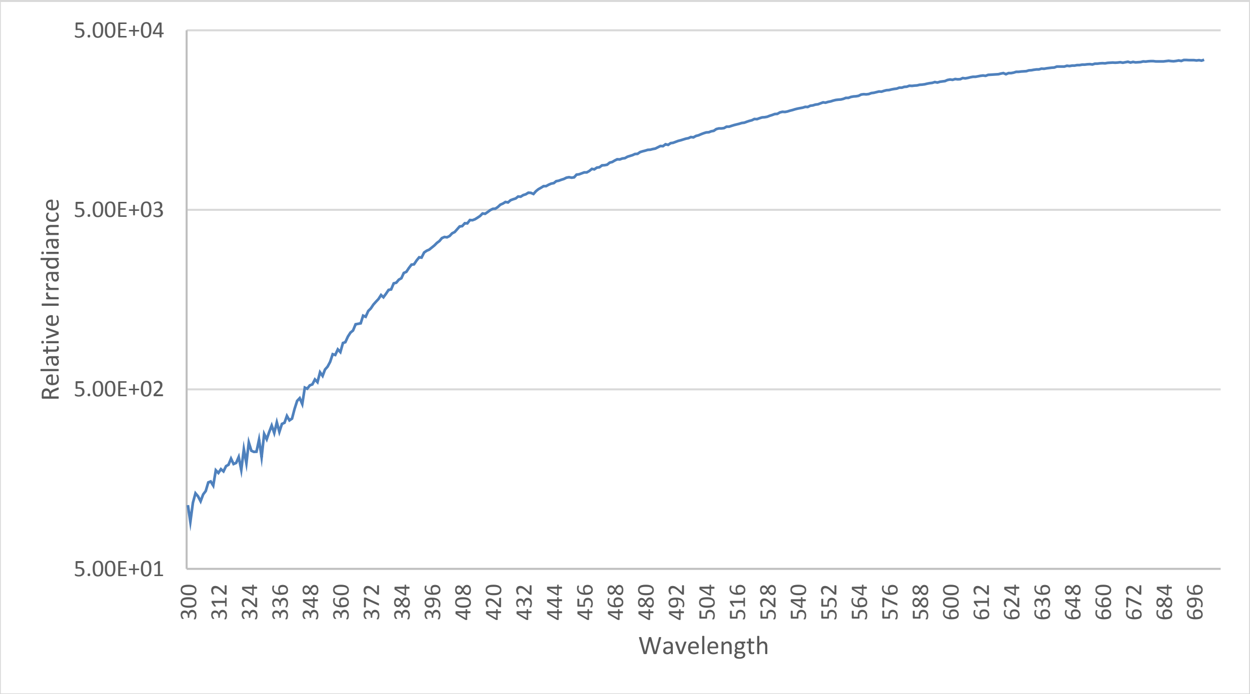
In sum, this issue should be dealt with to convince the reader that the colour calculations are indeed correct, and no hidden UV signal is responsible for the difference in discriminability of different colour pairs.

Response: We thank the reviewer for noting the potential that differences in UV could affect birds' discrimination abilities, and we have taken several steps to demonstrate that this possibility is unlikely. We re-measured the irradiance of our ambient lights to get a clean irradiance spectrum down to 300 nm and calculated the quantum catch of the UV cone. We found that the quantum catch of the UV cone represented only 1.5%-2.6% of the total quantum catch for each of our seven stimuli. We can therefore conclude that UV information available to the birds was minimal and very unlikely to affect our results. We include a detailed description of the methodology we used to come to this conclusion below.

First, to test whether our stimuli reflect any UV light, we re-measured the reflectance of our colour stimuli using a xenon UV-VIS light source coupled with a back-reflection fibre probe at an angle of incidence of 45 degrees. Previously, we had used the tungsten light source in our integrating sphere, given that the integrating sphere provides a measure of average reflectance from all angles of illumination. Below we include the new reflectance spectra for each of our seven stimuli. As can be seen in the figure, the reviewer was correct that our stimuli reflect light in the UV range (below 400nm). The critical issue, however, is not whether the stimuli reflect in the ultraviolet, but whether, once we have accounted for ambient viewing conditions, the UV cone of zebra finches is stimulated at a significant level as a result of this relatively low UV reflectance. As we show below, our ambient light source emitted very little light in the UV range, ultimately predicting very little activation of the UV cones.



As the reviewer noted, our previous ambient light irradiance spectrum was imperfect, as its values dropped below zero at low wavelengths (again due to the fact that a probe not properly calibrated in the ultraviolet range would have additional noise). We therefore re-measured irradiance spectra from two of the lights which we use during experiments, and took an average of these two spectra as our new ambient light irradiance spectrum.



We then re-calculated the quantum catch of zebra finch long, medium, short, and UV cones using our new ambient light spectrum and the original reflectance spectra from our stimuli. We first confirmed that the distances between our seven colours were unchanged from our previous calculations using the original ambient light irradiance spectrum that we presented:

Colour Step	1-2	2-3	3-4	4-5	5-6	6-7
Distance with old ambient light spectrum	4.1	2.9	2.3	3.8	3.4	3.4
Distance with new ambient light spectrum	4.2	3.0	2.3	3.8	3.5	3.4

This re-calculation led to only very minor changes in distance estimates, and there were no changes in the relative position of step distances: 1-2 was still the largest step, 4-5 the second largest, etc. We also note the reviewer's concern that one of our original ambient light spectra (included in the average of 18 light sources) was dimmer than the others. That light spectrum was included because it reflected real variation in one of the lights that we used when running experimental trials, and we felt it important to include all of the relevant data. However, we note that our two re-measured lights were of approximately equal brightness and we saw essentially no shift in the distances between colour steps.

Finally, using the reflectance spectra for our stimuli obtained from the integrating sphere, we calculated the percentage of total quantum catch (long+medium+short+UV) that was contributed by the UV cone for each colour. There was some variation between the 7 colours, but in each case the UV quantum catch constituted between 1.5 and 2.6 percent of total quantum catch.

Colour	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Percent UV Quantum Catch	2.6%	2.4%	1.9%	1.8%	1.6%	1.5%	1.5%

This outcome was unchanged if we instead used the reflectance spectra that we obtained using the xenon light source (range of UV catch was 0.8-2.1% of total quantum catch).

We therefore conclude that, although the Munsell colours we used do reflect some light in the UV range, it is still appropriate for the purposes of our experiment to use a trichromatic space excluding the UV quantum catch because (1) our stimuli have low reflectance in the UV range as compared to the visible range and (2) the ambient light used during our experiments (a halogen-tungsten light source) has almost no ultraviolet light in the UV range, thus resulting in UV quantum catch during our experiments being extremely low and thus likely having no impact on perception. Therefore, we do not believe that differences in UV information between colours can explain our results.

We have included a statement to the above effect in the main text on lines 104-109.

We have also made the following changes to the supplemental material:

1. We now provide zebra finch spectral sensitivity curves from 300-700 nm for UV, short, long, medium, and double cones.
2. We also provide the reflectance spectra obtained from the integrating sphere for our 7 colours from 300-700 nm.
3. The ambient light irradiance spectrum now reflects the average of our two re-measured lights.

2. Scaling of the model colour space
this is really a comment rather than criticism: making a difference between delta s and JNDs and defining one as 2,3 time the other is really only an academic exercise and irrelevant, as long as the real lever of receptor noise is not measured. You could equally well adjust the assumed noise level in the zebra finch photoreceptors by a factor of 2,3 and use 1 as the threshold. It may be important not to forget what model calculations are based on.

Response: We had not intended to define ΔS as 2.3 JNDs and thank the reviewer for pointing out that our language was unclear on this point. The Euclidean distances that we report here are derived from a colour space that is a representation of the Receptor Noise Limited Model (RNL) of colour discrimination, and are equal to the chromatic distance (ΔS) of the original RNL model. Some researchers equate these measures of ΔS with JND's. However, several studies have now demonstrated that a behavioural Just Noticeable Difference is often greater than one ΔS. We had intended to simply illustrate this point by citing as an example a published reference that 2.3 ΔS is a behaviourally-derived Just Noticeable Difference in honey bees, but we did not mean to imply that we used 2.3 as a scaling factor for zebra finches. To clarify this point, we have added text on lines 119-122 which summarizes the arguments we have just laid out, and which we hope will clarify the relationship between ΔS and JNDs.

3. Pass criterion

It would be good for a general reader to understand what "pass" really means in terms of statistics. So it would help to give what is the statistical probability that a bird chooses the two positive stimuli first, if she chooses by chance? the probability of the first correct choice is 1/3, and the second choice is 1/5 so in total, the chance to do both right should be 1/15 - correct? How often would that happen in 10 trials? Thus, what pass frequency is statistically different from chance? Why do trials in which the bird chooses one correct but a second incorrect disk have the same value as trials in which a bird chooses two incorrect ones? How would the data look like if, for instance, the % correct choices for each trial was taken for all 10 trials? This would allow a general reader to compare these data to the data obtained in other studies with more classical methods.

A simple case can illustrate my concern: think of a bird that makes one correct and one incorrect choice in each of 10 trials. If a bird makes 2 choices per trial, she would then have 10 correct choices out of 20. In your analysis, this bird would be counted as 0% pass. another bird makes 2 correct choices in 5 out of 10 trials, and no correct choice in the other 5 trials, adding up to the same number of correct choices, 10 out of 20. However, this bird would be counted a 50% passed trials. I am not saying both birds have the same performance, but I would argue that the difference may be more subtle than your analysis lets the reader think.

Response: We thank the reviewer for this careful analysis of our pass criterion, and have addressed this comment by re-performing our analysis with a less stringent pass criterion in which we considered a bird as passing if she flipped only a single bicolour disc before flipping any other discs. The results of this analysis are qualitatively identical to the results originally reported. Below we show the results from our model of pass frequency, with both the original (A) and this new pass criterion (B) presented side by side:

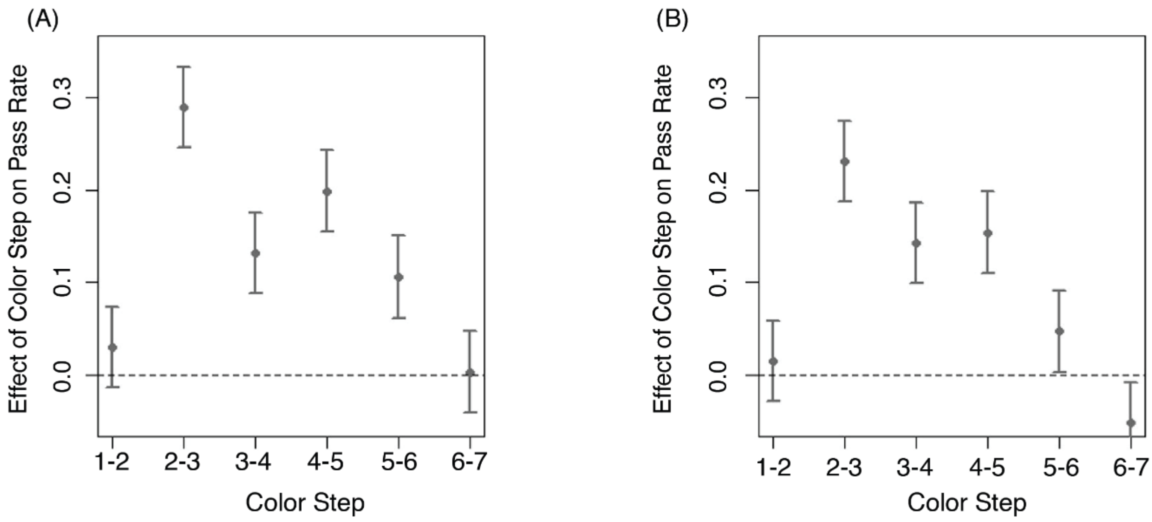


Fig S2. A comparison between (A) the model outputs presented in the main text and (B) model outputs if we use a more lenient pass criterion of a bird flipping a single bicolour chip before flipping any other chips. All results are qualitatively identical between the two models- we still observe a category boundary between colours 2 and 3, we still see no category boundary between colours 4 and 5, and the boundary in the blue-green range is still relatively weak as compared to the boundary in the orange-red range.

We have included this figure, as well as a table of the model output of these results in the supplement as Table S1 and Figure S2.

As the reviewer points out, the probability of a bird flipping the two bicolour discs first is 1/15, whereas the probability of choosing only a single bicolour disc before any others is 1/3. We have therefore kept the original pass criterion in the main text because choosing an outcome with a lower probability of occurring by chance gives us more statistical power and precision when comparing pass frequencies.

We now explicitly identify the probability of a bird passing the task by chance in the main text and we also direct readers to Table S1 and Figure S3 on lines 143-146.

4. Colour distances and colour steps

Even though there the authors make a correct statement that the best discrimination did not happen for the smaller colour steps, it would be very helpful to plot the percentage of correct choices as a function of the colour difference, not only over each colour step as if they were equal. Other colour discrimination studies often show very steep psychometric functions for discrimination.

Response: Please see our response to the reviewer's Point #5—we address Points 4 and 5 together below.

5. The contribution of each colour step to pass frequency
This is a rather complicated measure, which, again, does not take into account, the colour differences between stimuli (which, actually, don not sum up linearly if the colour loci are not on a line). Again, it would be helpful to present these data also in a way, in which the specific colour differences are taken into account.

Response: We appreciate the reviewer's points 4 and 5, and have taken steps to clarify this for readers. Specifically, we have added the requested figure (pass frequency vs. chromatic distance) in the supplement as Supplemental Figure S6. We have also added Supplemental Table S3, which displays the model outputs from a model that considers only chromatic distance as a predictor of pass frequency rather than individual colour steps, and compared the AIC scores of this model to the colour step model reported in the main text.

To aid the reviewer, we have also included all of this additional material here, along with a discussion of its implications for the reviewer's concerns.

Pass frequency does indeed increase as the distance between colour pairs increases (as shown in the figure below), but not as a steep (i.e. sigmoidal) psychometric function, as the reviewer reasonably suggested we might expect. The following figure shows the average pass frequency across birds for all colour comparisons versus chromatic distance, ΔS.

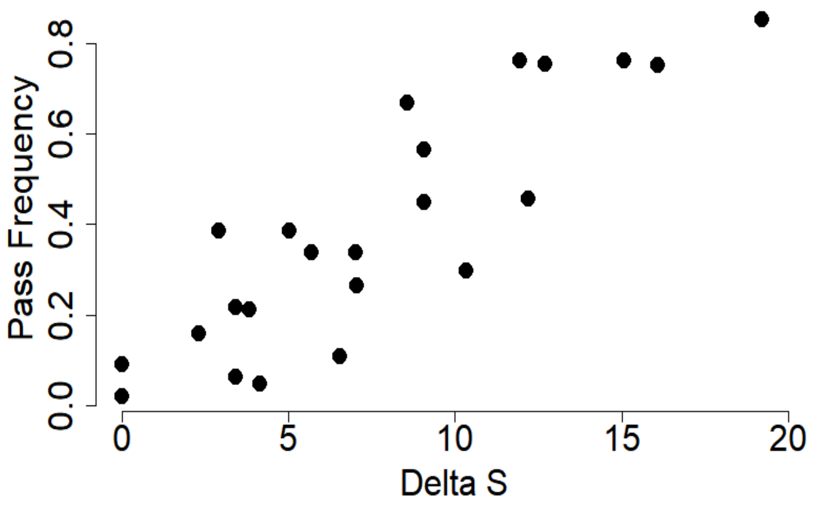


Figure S5. Average pass frequency across birds for each colour comparison versus chromatic distance, ΔS. Pass frequency increases as ΔS increases, but not as a sigmoidal psychometric function.

We built a mixed effects model that predicts pass frequency as a function of chromatic distance instead of as a function of the six binary colour steps. This model shows chromatic distance to be a significant predictor of pass frequency ($p<0.0001$;

Table S3). However, this distance-only model (AIC = -101) does not perform nearly as well as the model that we report in the main text (AIC=-244), which specifically considers the contribution of each colour step to pass frequency.

Table S3. Model outputs from a model that includes only a fixed effect of chromatic distance and random effects of bird ID*.

Parameter	Coefficient Estimate	Std. Error	P value
Intercept	0.05	0.03	
Chromatic Distance	0.044	0.002	<0.0001

*AIC of this model = -102. AIC of the colour-step model reported in Table # in the main text = -244

We would like to further address the reviewer's concern about the potentially un-observed role of chromatic distance by adding a fixed effect of distance to the colour-step model that we report in the main text (Table 1). However, because the colours that we chose are very close to falling in a straight line in chromatic space, this chromatic distance term is too tightly correlated with the binary fixed effect components of the model for such a model to be meaningful (i.e. the distance contained in a 1|4 comparison can be well approximated by adding the distances contained within the 1-2, 2-3, and 3-4 colour steps). Thus, we are prevented from building such a model. We hope this fact—that the distance between any two colours is nearly additive—satisfies the reviewer's concern that the distances between colour steps do not sum exactly because of their non-linearity in colour space.

In summary, we believe it is very unlikely that we lose important information by considering colour steps as our predictors rather than chromatic distance for three reasons. First, the colour-step model reported in the main text performs much better than a distance-only model. Second, pass frequency does not show a sigmoidal psychometric response to chromatic distance, suggesting that our treatment of unequally-sized colour steps as equal is very unlikely to affect our conclusions. And third, as we reported previously, the contribution of each colour step is not predicted at all by the chromatic distance contained within it (Fig S5).

6. Comparison with the previous study

A main difference between both studies was the presence or absence of achromatic differences between the colours. This should clearly be discussed before speculating about and referring to similarities to humans and connections to environments. Basically, the presence or absence of achromatic differences is the most parsimonious explanation that the differences in the strength of the results (if they hold based on a more classic evaluation of results in terms of statistically different choice frequencies in the different tests instead of pass frequencies that ignore single correct choices in trials, as outlined above).

> **Response:** We thank and agree with the reviewer. We have now placed the achromatic difference explanation first in our discussion, on lines 322-329.

Referee: 2

Comments to the Author(s)
The paper "categorical colour perception occurs in both signalling and non-signalling colours in songbird" is well written and asks an important question. In the previous paper the authors investigated colour categorisation along the red-orange direction in the colour space. This paper investigates colour categorisation along blue-green direction in the colour space.
The authors find that the birds trained to select two coloured disks demonstrate the best performance with the colours 2-3, which is at odds with the prediction based on colour distances and agrees with the hypothesis that there is a categorical boundary between colours 2 and 3. While it is plausible that birds indeed have categorical colour perception, the results can be explained using several other hypotheses. I do not think that alternative explanations can be ruled out. However alternatives need to be discussed.
An alternative to the colour categorisation hypothesis is the hypothesis of proximity judgements. The birds can simply select the colours that are more different (have larger colour distance) from each other. Usually colour distance is calculated on the basis of threshold measurements. The authors estimate colour distance using the receptor noise limited model. This is the simplest model that ignores the later stages of colour processing. Predictions of this model agrees reasonably with the results obtained using stimuli in the vicinity of achromatic point. However I remains uncertain if the model gives correct predictions for stimuli that the authors used. I suggest that the experimental results with birds are discussed in more details.
The distance based on thresholds does not necessarily agrees with perceptual colour distance. Birds also can select the most salient edge. In humans, edge detection is mediated solely by a luminance mechanism. The authors suggest that in birds it is mediated by double cones. However this hypothesis is based on limited experimental evidence. Possibly L and M cones can be involved in the process.

> **Response:** We thank the reviewer for carefully considering alternative explanations for our results. Much of the reviewer's concern stems from uncertainty regarding the accuracy of the RNL model for measuring distances between the colours that we used in this experiment. All calculations of chromatic distance for any colour study involve the use of a model, and it is of crucial importance to ensure that the model is trustworthy. In the case of our study, we have used the Receptor Noise-Limited (RNL) model to calculate ΔS, a measure of the chromatic distance between colours. While the RNL model is well-accepted, we agree with the reviewer that the RNL model has limitations. It purports to predict colour-mediated behaviour as a result of receptor-level mechanisms, and our results show that it does not always succeed. Yet, we do not

interpret this outcome to be a failure of the model to accurately predict quantum catch or receptor level processes.

The reviewer is correct that the RNL model is agnostic as to processing of inputs above the photoreceptor level. We see this simplicity as a strength, not a weakness of using the RNL model when asking questions of categorical perception. We think it is very likely that categorical perception is the result of processing beyond the photoreceptor, although how exactly this processing occurs remains an open question. Our goal here is only to show that birds respond categorically to stimuli that, based on the best available knowledge and models, vary continuously in a colour space dictated by quantum catches (the relative stimulation of each photoreceptor type in response to a given colour).

We have elaborated on the role of processing of visual inputs above the level of the photoreceptor and the potential for more complex models in the discussion on lines 355-359.

With regards to the reviewer's comments about edge effects, we do not state in our manuscript that we believe edge effects to be the result of differences in luminance, so we are unsure how best to respond. We agree with the notion that edge effects may play a role in birds' ability to discriminate colours, but we do not believe edge effects contribute to our results. Rather, we believe that by placing two colours on a single disc with an edge between them, we have maximized the chance that birds are able to discriminate between the two colours. Yet, we still observe categorical perception along the blue-green range under circumstances where discrimination within a category should be easiest. We argue that this outcome strengthens, rather than weakens, our conclusions.

Preference for colour combination also may explain the results.

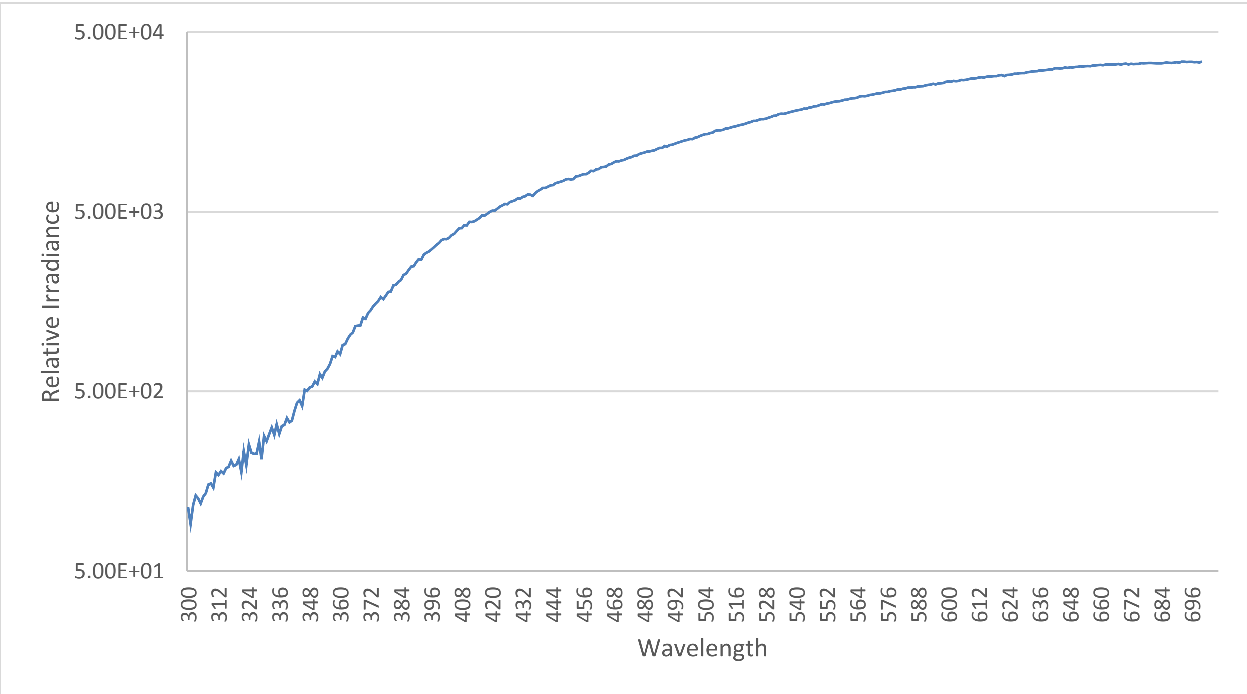
Response: We do not agree that preference for bicolour discs could explain our results. Indeed, we trained birds to specifically prefer bicolour discs. We then asked birds to demonstrate this trained preference by identifying bicolour discs whenever they were able. Experimental trials then allowed birds to identify those colour comparisons they were able to perceive as colour combinations and those that they perceived as solid colours.

Additionally, during the training phase of our protocol, we initially rewarded birds for flipping all discs, including solid blue, solid green, and bicolour discs. In these training stages, birds flipped discs essentially at random, and we did not observe any systematic avoidance or preference for either blue, green, or bicolour discs. Therefore, we do not think the birds had any innate preference or dislike of either blue or green that would have biased our results. Indeed, in the majority of training trials, birds consistently flipped all six discs, at first in random order, without displaying any obvious preference for any one type of disc.

Minor points

1. The authors state that as ambient light they used illumination A. However experiments were performed under halogen bulbs with similar colour temperature but spiky spectrum. For humans, these two illuminations are almost metameric. However, for birds, quantum catches calculated under these illuminations must differ substantially. Please, calculate quantum catches using correct illumination. The colour distances may change substantially and therefore the conclusions may be affected. It may also be useful to measure the spectrum of light reflected from Munsell chips illuminated by halogen bulbs.

Response: We agree with the reviewer that it is important that we demonstrate that the distances between colour distances are consistent with different illuminations. To attempt to address the spikiness the reviewer refers to, we have re-measured the ambient light present in the bird room by taking relative irradiance measures from two bulbs and taking steps to reduce the noise in the 300-450 range of the spectrum (see response to Reviewer 1). Our revised ambient irradiance spectrum is as follows:



The reviewer is also correct that we calculated quantum catches for our colour stimuli using illuminant A, but performed our experiment halogen illumination. The reasons for doing this are that the halogen bulbs we used have an irradiance spectrum that is quite similar to the irradiance spectrum of illuminant A (a standard illuminant meant to represent a halogen-tungsten light source), and using a standard illuminant like Illuminant A makes our calculations easier for other researchers to repeat.

To ensure that our use of Illuminant A in our calculations, but halogen lights in our experiments, have not biased our results, we have re-calculated quantum catches for each colour using both our revised ambient light irradiance spectrum and an illuminant A irradiance spectrum. The following table shows the calculated distances between colours (in units of ΔS) using three different illuminants.

Colour Step	**1-2**	**2-3**	**3-4**	**4-5**	**5-6**	**6-7**
Distance with old ambient irradiance spectrum	4.1	2.9	2.3	3.8	3.4	3.4
Distance with new ambient irradiance spectrum	4.2	3.0	2.3	3.8	3.5	3.4
Distance with Illuminant A irradiance	4.1	2.8	2.3	3.7	3.4	3.4

This re-calculation led to only very minor changes in distance estimates, and there were no changes in the relative position of step distances: 1-2 was still the largest step, 4-5 the second largest, etc.

Thus, recalculating our colour distances using an illuminant A irradiance spectrum does not have an impact on our results or conclusions.

At the editor's discretion, we are happy to include some version of the above table as supplemental material.

2. Please indicate in the figure 1 the numbers corresponding to each colour. The Munsell name is less important, you can add it the figure legend.

Response: We have made the requested change and now indicate the numbers corresponding to each colour in the actual figure, and have added the Munsell names to the figure legend.

Appendix B

Reviewer(s)' Comments to Author:

Referee: 1

Comments to the Author(s).
Congratulations! The new version of the manuscript the authors have taken comments into account, and clarified a lot of uncertainties present in the original submission.

We thank the reviewer for their useful feedback in the first round of revisions- their input has made the manuscript much stronger.

The authors present new measurements, have calculated their pass criterion in an additional way and show that the result is not changed. However, a lot of the added information can only be seen in the reply to the reviewers, or if the reader plots curves from the data in the supplement. I therefore suggest to add the plots (e.g. reflectances of the colour stimuli, light measurements and % quantum catch of the UV cone) to the supplementary text file. It costs nothing but helps the reader.

We have added the following additional plots to the supplemental excel file:

- Reflectance spectra of the stimuli (from the integrating sphere)
- The irradiance spectrum from the experimental lighting conditions
- Bar plots of the total quantum catch across each cone type for each colour

In the spreadsheet, please give the units for each measurement. I doubt it is just "photons" for the light, and it should say percent reflectance for the Munsell chips and also how the cone sensitivity curves are normalized.

The reviewer is correct- the measure is really the relative number of photons compared to other wavelengths. We have therefore revised our unit to read, "normalized quantal irradiance" which is unitless.

We now label our reflectance data as "% reflectance by color." We have also altered our reference to the source of the zebra finch sensitivity curves (line 103) to make clear that they were normalized by others (specifically Lind, 2016), not by us.

In the supplement file, abbreviations should be explained. I guess most people know that CI is supposed to mean confidence interval, but it should be said at first mention anyway. Maybe in 50 years from now, CI typically is used for something completely different.

We have replaced "CI" with "confidence interval" in the supplement.

I specifically appreciate the additional analysis of choices using a second, more relaxed choice criterion. My original point was not that the authors should change their criterion. The point was rather that that the reader should be convinced that this choice criterion is reliable, and you show it is – great. Introducing new methods is most helpful if the results are still comparable with those of other studies.

The same applies to the "classic" analysis of the results. What it shows is that the results are much more spread than they are in a classical discrimination threshold test. If jnd calculations are in an acceptable range (no severe over- or under-estimation of noise) then a bird should definitely be at 80% correct choices with 10 jnds – if the bird really does its best.

That this does not always happen immediately leads the reader to think about the reasons and very likely the reason is that the experimental design is such that birds to do show their absolute best discrimination ability. That is fine, as it likely is the condition under which they show signa of using categorization instead. However, I think it is worth mentioning this fact. This may become a philosophical discussion but IF categorical colour perception means that categorization overrides physiology (by which mechanism ever) THEN the conditions under which this happens need to be defined.

We appreciate the reviewer's feedback on this topic. We have added a sentence to our discussion (lines 358-359 and 365-369) that specifically states that there are particular conditions under which some kind of processing above the photoreceptor level becomes relevant to colour perception. Specifically, these conditions occur when the comparison being made is below some threshold of perceptual distance, (which seems to correspond to a ΔS of about 11 in our case). When the distance between two colours falls below such a threshold, then categorical perception can occur.

Please mention what Table S3 shows, in your results section. You only refer to it, but don't mention the important fact that the colour step model does explain the data better than the model just taking chromatic distance into account. In the end, this is the most solid statistical proof of the categorization so should clearly be written out as it is in the reply to the reviewers. As with other comments to the first version: most of them were thought to ask you for better prove that your interpretation of data is correct. Strong claims require strict and critical tests, and it is great to see that you can give these, so please present them to the readers! The point with reviewers is that they should be the most critical readers, and play the role of advocati diaboli, to make sure the paper convinces other critical readers as well!

We agree with the reviewer and have added a sentence explicitly stating that the colour step model explains the data better than the model which only takes chromatic distance into account on line 249-251.

In reply to the previous comment on jnd versus ΔS: As I said before, this was just a comment, and I don't request you to do anything about this but my main point was that this is still an academic discussion as the noise levels in zebra finch receptors have never been measured so nobody knows the "real" absolute jnds for the species. Thus we do not know whether the best behavioural thresholds are higher than these absolute thresholds because we do not know the absolute thresholds.

The most simple suggestion to resolve this and the point made before, could be that in the task used in this experiment, birds do not really perform at their best but the only way to find this out would be do give them the chance to perform as good as they can and that would imply a rather boring threshold experiment. For the present paper, there is need to do this.

We thank the reviewer and agree that follow up experiments as a part of a separate study should be performed to consider exactly what conditions categorical perception is most likely to occur.

In short, my suggestion is that
1. the authors add more of the graphs that they present in the reply to the reviewers, in the text supplement, and that they check that they explain all abbreviations and correct units in texts and tables.
2. The authors should mention that discrimination – according to colour distances – is actually worse in this experiment than would be expected in an experiment that has the goal to find the best discrimination with lowest threshold. It is under these conditions, that categories play a role.

We have addressed both of these main suggestions from the reviewer, as described in above responses.

Else, this is now a very exciting manuscript, specifically as it now gives more method-critical details which indeed can convince even a critical reader! I also find the reply to the comments made by the second reviewer very convincing.

Again, we are grateful for the reviewer's care and thorough consideration of our work. The reviewer served their role perfectly and their attention has greatly strengthened our manuscript.

Referee: 2

Comments to the Author(s).
I have two major concerns with the revision.

1. In my previous review, I pointed that the authors calculated quantum catches using incorrect illumination. However, the authors did not change all results using correct illumination. They argue that results change only marginally if the wrong illumination is used. As long as the calculations of quantum catches are not reliable all conclusions are not valid.

We disagree with the reviewer that we used incorrect illumination, either when measuring the reflectance spectra of stimuli or during trials. We report quantum catches calculated using illuminant A (a standard illuminant spectrum) as the illuminant in order to improve replicability and generalizability of our results- not all halogen lights are identical and these differences might lead to small changes in quantum catch calculations. However, we showed in our previous response that the experimental lighting that we used yielded nearly identical estimates of perceptual distances between stimuli using the receptor noise limited model. Therefore, we disagree that there is a “wrong” and a “right” illumination; rather, there is the experimental illumination, which is essentially identical to the standard illuminant used in calculations. However, using either does not change our results nor our interpretation of these results.

In the response to reviewers, the authors present the spectrum in logarithmic scale. This scale is not relevant to calculating quantum catches.

Our inclusion of the logarithmic scale in the previous response to reviewers was not meant to imply that we used logarithmic data to calculate quantum catch. Instead, we provided it as a conservative illustration that would maximize any spikiness or discontinuities in our spectra, in order to be as transparent as possible to readers. It is the absence of any such spikes, even at a logarithmic scale, that points to the quality of the spectra that we were able to obtain.

I am surprised that the spectrum does not contain spikes around 600 nm. These spikes are likely to result in non-marginal changes in the calculated L cone quantum catch.

One of the great benefits of using halogen lights is that they are not spiky at the longer visible wavelengths. Because we do not have any spikes in irradiance, we do not believe that there are any unaccounted for errors in our calculation of L cone quantum catch.

In the supplementary materials, the units of illumination are not given. Which units have been used to calculate quantum catches?

Our ambient light spectrum data are unitless, as they show normalized quantal irradiance.

How was spectrometer calibrated? What was the sampling rate and spectral resolution? How were the 1nm interval data generated?

The spectrometer was calibrated using a tungsten light source with a color temperature of 3100K. The sample rate was 1000 full spectra per second (we used 3000 measurements per spectrum, so our spectra was the result of 3s of sampling time) and the spectral resolution was 0.1nm. We took the average of three measures taken per nanometer to generate the 1nm interval data. We are not familiar with examples of these metrics being included in methods sections, and so have not included them at this time, but we are willing to include this information at the editor’s discretion.

I suggest to check if the light above 700nm contribute to quantum catch of L cone, because the lamp may have high output in the red- near infra-red part of the spectrum.

Spectral sensitivity curves for zebra finches are available only up to 700nm, but we believe that it is very unlikely that sensitivity to wavelengths >700nm would affect quantum catch calculations. Spectral sensitivity of zebra finches is nearly zero at 700 nm (~1.3% of maximum sensitivity, which occurs at 607 nm), and the stimuli that we use had very little reflectance in this range (<8% of peak reflectance).

To convince a reader that there is no mistake in calculations or measurements, more information must be given in the supplementary materials.

We now include the following additional figures in the supplemental excel file:

- Reflectance spectra of the stimuli (from the integrating sphere)
- The irradiance spectrum from the experimental lighting conditions
- A box plot of the total quantum catch across each cone type

I also suggest that the authors reanalyse the data from previous paper using correct illumination.

We refer the reviewer to our above response to the previous comment about the illumination that we used both in our experiments and to calculate chromatic distances.

2. In my previous review, I suggested that the authors discuss alternative explanations. However the authors preferred not to do it. Instead, they wrote a lengthy explanations addressed to reviewer.
All conclusions are based on the deviations of bird performance from that predicted by the model. This must be clearly stated in the paper. I find this line of arguments to be thin and I am not entirely convinced by the authors' arguments.

We have added a sentence to this effect on lines 359-361 where we write that, "Our conclusions about categorical perception hinge on differences between measured behavioural responses to colour stimuli and predicted chromatic distance from the RNL model, which only considers visual inputs at the photoreceptor level"

Multiple reasons that may explain the deviation of model predictions from experimental results. Readers must be convinced that the deviations cannot be explained by errors in calculations or measurements (see comment 1, sufficient information must be given in the Supplement). Also, the alternative explanations must be discussed in the paper, and readers may decide if the arguments in favour of the hypothesis of categorisation are convincing.

We note on lines 359-361 that what we are detecting are differences in discrimination behavior from what would be expected by the RNL model. We suggest that these differences are the result of processing above the photoreceptor level, and in particular the pattern in which the results deviate from the predictions of the RNL model are precisely consistent with the predictions of categorical perception.

Minor points:

Line 101 wrong format of Eq.

We thank the reviewer for pointing out this error and we have corrected it.

Spectral sensitivities of zebra finch. Lind did not measure the sensitivities, leave this reference for modelling and give reference to msp and oil droplet modelling.

We have added a reference on line 103 to Bowmaker et al (1997), "Visual pigments and oil droplets from six classes of photoreceptor in the retinas of birds," in which the spectral sensitivity of the zebra finch was first quantified. We also retain the Lind reference, because Lind's published sensitivity curves account for oil droplet absorbance and transmission through zebra finch ocular media.