THE ROYAL SOCIETY PUBLISHING

PROCEEDINGS B

Tolerant and intolerant macaques show different levels of structural complexity in their vocal communication

Nancy Rebout, Arianna De Marco, Jean-Christophe Lone, Andrea Sanna, Roberto Cozzolino, Jérôme Micheletta, Elisabeth H. M. Sterck, Jan A. M. Langermans, Alban Lemasson and Bernard Thierry

Article citation details
Proc. R. Soc. B **287**: 20200439.
http://dx.doi.org/10.1098/rspb.2020.0439

Review timeline
Original submission: 27 February 2020
Revised submission: 13 May 2020
Final acceptance: 15 May 2020

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

Review History

RSPB-2020-0439.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?
Excellent

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?
Yes

Is it clear?
Yes

Is it adequate?
Yes

Do you have any ethical concerns with this paper?
No

Comments to the Author
Review of RSPB-2020-0439 macaque vocal / social complexity

The authors compare four species of macaque in terms of individuals' levels of tolerance and vocal complexity in the species. The authors had a nice small comparative data set here in that one pair of macaques (Tonkean and crested) are tolerant whereas the other pair (rhesus and Japanese) are intolerant. The authors find that in general, the tolerant species exhibit greater complexity in major contexts of vocal signaling than the intolerant species.

This is a well-written manuscript and described a solid (though small in terms of number of species) comparative study. Some important strengths of the work include the breakdown of vocal signals into different contexts (affiliative, aggressive, neutral) and the assessment of vocal complexity in finer detail than most studies carry out. I do have a number of concerns about the work or the reporting of the work, however.

First, from the onset it was not clear why or how the phylogenetic hypothesis makes a different prediction regarding vocal complexity in comparison to the social complexity hypothesis, since it appears from the text that the two tolerant species are most closely related (of these four species) – see lines 131-137.

Second, it was not clear to me why the authors would claim that the social complexity hypothesis would predict greater complexity in tolerant species for aggressive signals only, and not affiliative (I can see why differences would not be predicted for neutral signals). Individuals are still needing to manage the behavior of others through signaling in affiliative contexts, so it seems to me one would expect socially complex species to exhibit greater complexity in signals in those contexts as well as in aggressive / agonistic contexts, under the social complexity hypothesis. Gustison et al. (2012) found, for example, that geladas were similar to other and less socially complex species in terms of their vocal signals except for the greater complexity of their affiliative and reconciliatory vocalizations. See:
• Gustison, M. L., A. le Roux, et al. (2012). "Derived vocalizations of geladas (Theropithecus gelada) and the evolution of vocal complexity in primates." Philosophical Transactions of the Royal Society B-Biological Sciences 367(1597): 1847-1859.
The authors seem later to admit as much in terms of the complexity of affiliative contexts, in lines

425-432.

Third, it is crucial to report the extent of recording work across dates and times for each species, rather than just in terms of sample duration when data were collected (see lines 180-187). I doubt this is the case in this study, but a worry would be that sampling effort across dates for rhesus and Japanese macaques, as opposed to Tonkean and crested macaques. If, for example, rhesus and Japanese were recorded in a relatively short time span, that could result in diminished contextual and motivational variation, contributing to decreased variation in signal structures.

Fourth, the authors need to carry out inter-observer reliability measures to support the coding of behavior categories here. Hopefully they are able to assess reliability for coding of signals produced in the three different contexts (lines 205-210), though if that were done live, perhaps there is no way to assess reliability. Nonetheless, inter-observer reliability could easily be done for the measure of vocal diversity (lines 270-274).

Additional comments by line number:

34. Perhaps "within" is better here than "between"? Or do the authors mean "among" instead?

50. By "structural" do the authors mean the term in the way Hinde (1976, in journal Man) meant it?

77. Instead of "undermines" here, do the authors really mean "limits"?

90. "The ability to produce uncertainty" seems an odd way of phrasing the idea for "Greater diversity among and within units of a system". The "to produce uncertainty" implies either an agency or a function that I am not sure the authors intend.

105-107. I don't think these two ideas perfectly align with one another, but wonder if a counter-argument to what the authors write here is that there would seem to be a great deal of honesty / reliability guarantees in signaling even in contest behavior – see the Bradbury & Vehrencamp chapters on this in their textbook or Searcy & Nowicki's book on the evolution of animal communication. If so, one might not necessarily expect much in the way of uncertainty in contest situations.

108-111. I understand the argument here for neutral situations, but don't understand it at all for affiliative situations. The authors should really develop this more as to why they hold this view, or consider the point above (and the Gustison 2012 reference) that affiliative situations, too, should result in greater uncertainty. In either case, a bit more discussion about the view regarding affiliative contexts would help a reader here.

143. Should be "Social Complexity hypothesis"

144-147. The authors address this later in the discussion, but they communicate a directionality (social -> vocal) here that their data cannot address. It is equally likely in this study that vocal -> social and that vocally complex species (who evolved that complexity presumably for other reasons) are the only ones able to evolve complex social structures – this was the basic interpretation of the McComb & Semple comparative study the authors cite (although that study, too, could not determine directionality!).

252. Perhaps the authors could add a sentence as to why this cut-off of 0.7 is advocated in the study they cite? It seems rather arbitrary.

262. I am not entirely sure what "stabilize cluster results" means – could the authors describe this in a sentence or two here?

395-407. I kept reading and rereading this section in comparison to the text in lines 390-394 and Fig 2, and still do not understand how the authors are making the interpretations they make here. From the results and Fig 2, it looks as though the differences seen between tolerant and intolerant species in terms of aggressive contexts are the same differences seen as for affiliative contexts, though perhaps not in as "statistically" clean a way as for aggressive contexts.

414-416. This sentence seems like a truism, rather than an "It could be . . ."

421-425. I am on the "information" side of the "information / manipulation" debate, and so had a bit of a laugh when the authors followed a sentence referencing Morton with a sentence describing information in signals – a nice touch! And, for me, a perfectly appropriate touch, though I am not sure Morton and colleagues would see it that way.

Finally, I realize a lot of what I raise above is critical, but I want to end on a positive note to circle back to my opening comments. I think this manuscript and study could make a major contribution to the literature on communication and what factors influence vocal complexity in species. The authors did a huge amount of work here, tackling an important question with sophisticated approaches to vocal complexity. The manuscript and study are already quite strong – I hope my comments, questions, and concerns are helpful in making the manuscript even stronger.

Review form: Reviewer 2

Recommendation
Accept with minor revision (please list in comments)

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

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

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

Is the length of the paper justified?
Yes

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

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
I like the comparative approach in this study. I think it is very useful to compare the use of vocalizations across closely related species in this way. The data sets are impressive and the statistical analyses seem sophisticated and sound. The differences are interesting and worthy of discussion. My main concern is with the overall presentation. It seems to try a bit too hard to present things as fitting with the social complexity hypothesis for communicative complexity. I would prefer to see the results presented in a bit more exploratory way and the fit with the complexity hypothesis discussed as one possibility. These are interesting results but I am not sure we can say so definitively what they mean.

Here are a few reasons why I don't see the data fitting with the hypothesis as cleanly as is presented. First, it is not clear what the diversity results really indicate. Looking at Table 4 shows that numbers of different call types used in the agonistic context are the same for Japenese+rhesus vs Tonkean+crested (only J+r use food calls and growls, only T+c use screech and soft grunt, the rest are shared). Also, the 3D figures indicate the J+r species have bigger clusters despite having fewer clusters. It seems that one could just as easily emphasize these features to indicate similarity in diversity. Also the affiliative calls show similar clustering differences as the agonistic clustering results (although slightly less clean) which doesn't support the prediction.

Furthermore, I don't quite agree with the use of graded calls as a measure of vocal flexibility and vocal complexity. I think the issue might lie with using potential information instead of realized information. With a coin flip, a balanced coin has more potential information than a biased coin because there is more uncertainty that can be resolved when the result is revealed. But it is the revealing of heads or tails that provides the information, not the ambiguity. The scoring of entropy used here is a measure of how difficult it is to categorize the calls, or the likelihood that calls fall acoustically between clusters. It is certainly possible that the presence of intermediate calls could indicate more potential information transfer but only if these calls communicated something different from what is communicated by the calls closer to the clusters. This would, essentially, indicate the presence of more call types. But that is not usually what happens with graded signals--typically the receiver uses contextual cues to place the intermediate calls in one of the clusters ("its heads"). The fact that the receiver has to do extra work to decode the signal seems consistent with the signal containing less, rather than more, information. In animal communication, information is usually used to refer to the reduction in uncertainty, not the uncertainty itself. An ambiguous signal reduces the uncertainty to a lesser extent. And intermediate signals (that have to be assigned by the receiver to one of the clusters) are ambiguous. I realized that these are difficult and contentious issues. I think it is certainly interesting that the degree of call gradation differs and this is worth discussing, I just think the issue could be discussed more carefully.

Review form: Reviewer 3 (Kirsty Graham)

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?
Yes

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

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

Is it accessible?
Yes

Is it clear?
Yes

Is it adequate?
Yes

Do you have any ethical concerns with this paper?
No

Comments to the Author
I enjoyed this paper and found it very clear to read. There was enough detail in the methods for it to be entirely replicable, from details about the study groups to equipment to analyses. My main question is around the handling of the affiliative context – it seems that the tolerant species also had more vocal flexibility and diversity in the affiliative context, but it's not really included in the discussion. It seems unsurprising that a species with more uncertainty in aggressive outcomes would also have more vocal complexity in affiliative contexts, as these are also contexts in which the individuals are managing their social standing/relationships. Perhaps the authors could either explain why they didn't consider the results from the affiliative context to be compelling, or could put more detail into discussing those results.

Introduction: Overall, the introduction gives nice, clear descriptions that provides readers with the background needed to get the most out of this paper. The descriptions of macaque social behaviour and structures are well detailed, and supports the justification for comparisons within this genus. There were just two points where more detail would be useful: (1) the definition of structural factors, and (2) information theory.

Line 56: the definition of "structural factors" is not very clear. It will be difficult for readers who are less familiar with environmental vs structural factors.

Line 71: the following paragraphs do a good job of outlining the limitations of how we currently measure social and vocal complexity.

Line 91: should this be "a means" (grammatical question)? Also, please explain how information theory helps us to measure uncertainty. This statement makes it seem obvious, but if readers aren't familiar with information theory they won't follow this point.

Line 140: It took me a moment to parse how hypothesis 2 and 3 are different. I would suggest re-ordering Hypothesis 2 – " Greater similarity should occur in more closely related species, for any variable, and regardless of the social context, so we should find more differences between Tonkean and crested macaques on the one hand, and Japanese and rhesus macaques on the other, than within each of these species pairs across variables and contexts."

Methods:
Very clear descriptions of the facilities/habitats and group compositions! I have reviewed a couple of papers recently that were lacking this, so thanks for including! The methods are detailed to the degree that they are almost completely replicable. This is unusual and very welcome.

Line 202 : "Contexts were defined according to the behaviours that could occur in the 3 s before or after" It matters whether the context is before or after the call – can you clarify what you mean by "3s before or after"? What if one context happens in the 3s before and another context happens in the 3s after?

Line 270: "We can measure vocal diversity by the number of call types in the repertoire of a species" Unclear whether you are taking the repertoire from this reference or are using the method described in this paragraph. If the former, should report species' repertoire sizes here, and also would be good to quantify if the groups used all of their species' repertoire. If the latter, should clarify that you are doing this "as follows".

Discussion:

Line 389: In the results, you do report a significant difference in vocal diversity in the affiliative context as well – why are the differences that you found for the affiliative context not as important as the ones from the agonistic context? Also on line 412 you say that "No more patterns were uncovered regarding vocal diversity" – while on line 336 your results show that rhesus macaques had smaller number of clusters than the other species and that rhesus macaques had a smaller number of clusters than the Tonkean/crested pair. This seems to reflect the pattern for aggressive behaviour, although perhaps not as strongly.

It seems that for vocal diversity and vocal flexibility there is evidence that the tolerant species are higher than the non-tolerant species. I don't see why this is a problem for the social complexity hypothesis: a species that has less certainty may need more affiliative strategies as well as more agonistic ones. On Line 414 you briefly posit this, and it seems quite unremarkable to say that affiliative interactions are more important than neutral ones. It feels like affiliative behaviours are being dismissed, because the focus was on aggressive behaviours, and that the role of affiliative behaviours in uncertain social systems is glossed over. Line 427 goes some way to expressing the importance of affiliative behaviours, so I wonder if it is more the way the discussion has been structured than the authors' actual interpretations?

Overall, I find this paper very compelling and I hope that my review helps the authors to bring out their key findings and interpretations. The final two paragraphs gave the paper a nice conclusion.

Decision letter (RSPB-2020-0439.R0)

24-Apr-2020

Dear Dr Rebout:

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,
Dr Robert Barton
mailto: proceedingsb@royalsociety.org

Associate Editor
Board Member: 1
Comments to Author:
All reviewers were very favorable in their appraisal of your article and I agree that it is comprehensive and interesting study. The comparative approach is particularly impressive. However, the reviewers also note a number of areas of concern that require further consideration. One common theme was your treatment of calls under affiliative vs agonistic contexts and I agree that your rationale here and interpretation of the data need clarifying. The reviewers offer a number of other points for review that I will not reiterate here but I do encourage you to carefully consider each comment and question provided both in terms of the framing of your hypotheses and the treatment and interpretation of your data.

Board Member: 2
Comments to Author(s):
(There are no comments.)

Reviewer(s)' Comments to Author:

Referee: 1

Comments to the Author(s)
Review of RSPB-2020-0439 macaque vocal / social complexity

The authors compare four species of macaque in terms of individuals' levels of tolerance and vocal complexity in the species. The authors had a nice small comparative data set here in that one pair of macaques (Tonkean and crested) are tolerant whereas the other pair (rhesus and Japanese) are intolerant. The authors find that in general, the tolerant species exhibit greater complexity in major contexts of vocal signaling than the intolerant species.

This is a well-written manuscript and described a solid (though small in terms of number of species) comparative study. Some important strengths of the work include the breakdown of vocal signals into different contexts (affiliative, aggressive, neutral) and the assessment of vocal complexity in finer detail than most studies carry out. I do have a number of concerns about the work or the reporting of the work, however.

First, from the onset it was not clear why or how the phylogenetic hypothesis makes a different prediction regarding vocal complexity in comparison to the social complexity hypothesis, since it appears from the text that the two tolerant species are most closely related (of these four species) – see lines 131-137.

Second, it was not clear to me why the authors would claim that the social complexity hypothesis would predict greater complexity in tolerant species for aggressive signals only, and not affiliative (I can see why differences would not be predicted for neutral signals). Individuals are still needing to manage the behavior of others through signaling in affiliative contexts, so it seems to me one would expect socially complex species to exhibit greater complexity in signals in those contexts as well as in aggressive / agonistic contexts, under the social complexity hypothesis. Gustison et al. (2012) found, for example, that geladas were similar to other and less socially complex species in terms of their vocal signals except for the greater complexity of their affiliative and reconciliatory vocalizations. See:
• Gustison, M. L., A. le Roux, et al. (2012). "Derived vocalizations of geladas (Theropithecus gelada) and the evolution of vocal complexity in primates." Philosophical Transactions of the Royal Society B-Biological Sciences 367(1597): 1847-1859.
The authors seem later to admit as much in terms of the complexity of affiliative contexts, in lines 425-432.

Third, it is crucial to report the extent of recording work across dates and times for each species, rather than just in terms of sample duration when data were collected (see lines 180-187). I doubt this is the case in this study, but a worry would be that sampling effort across dates for rhesus and Japanese macaques, as opposed to Tonkean and crested macaques. If, for example, rhesus and Japanese were recorded in a relatively short time span, that could result in diminished contextual and motivational variation, contributing to decreased variation in signal structures.

Fourth, the authors need to carry out inter-observer reliability measures to support the coding of behavior categories here. Hopefully they are able to assess reliability for coding of signals produced in the three different contexts (lines 205-210), though if that were done live, perhaps there is no way to assess reliability. Nonetheless, inter-observer reliability could easily be done for the measure of vocal diversity (lines 270-274).

Additional comments by line number:

34. Perhaps "within" is better here than "between"? Or do the authors mean "among" instead?

50. By "structural" do the authors mean the term in the way Hinde (1976, in journal Man) meant it?

77. Instead of "undermines" here, do the authors really mean "limits"?

90. "The ability to produce uncertainty" seems an odd way of phrasing the idea for "Greater diversity among and within units of a system". The "to produce uncertainty" implies either an agency or a function that I am not sure the authors intend.

105-107. I don't think these two ideas perfectly align with one another, but wonder if a counter-argument to what the authors write here is that there would seem to be a great deal of honesty / reliability guarantees in signaling even in contest behavior – see the Bradbury & Vehrencamp chapters on this in their textbook or Searcy & Nowicki's book on the evolution of animal communication. If so, one might not necessarily expect much in the way of uncertainty in contest situations.

108-111. I understand the argument here for neutral situations, but don't understand it at all for affiliative situations. The authors should really develop this more as to why they hold this view, or consider the point above (and the Gustison 2012 reference) that affiliative situations, too, should result in greater uncertainty. In either case, a bit more discussion about the view regarding affiliative contexts would help a reader here.

143. Should be "Social Complexity hypothesis"

144-147. The authors address this later in the discussion, but they communicate a directionality (social -> vocal) here that their data cannot address. It is equally likely in this study that vocal -> social and that vocally complex species (who evolved that complexity presumably for other reasons) are the only ones able to evolve complex social structures – this was the basic interpretation of the McComb & Semple comparative study the authors cite (although that study, too, could not determine directionality!).

252. Perhaps the authors could add a sentence as to why this cut-off of 0.7 is advocated in the study they cite? It seems rather arbitrary.

262. I am not entirely sure what "stabilize cluster results" means – could the authors describe this in a sentence or two here?

395-407. I kept reading and rereading this section in comparison to the text in lines 390-394 and Fig 2, and still do not understand how the authors are making the interpretations they make here. From the results and Fig 2, it looks as though the differences seen between tolerant and intolerant species in terms of aggressive contexts are the same differences seen as for affiliative contexts, though perhaps not in as "statistically" clean a way as for aggressive contexts.

414-416. This sentence seems like a truism, rather than an "It could be . . ."

421-425. I am on the "information" side of the "information / manipulation" debate, and so had a bit of a laugh when the authors followed a sentence referencing Morton with a sentence describing information in signals – a nice touch! And, for me, a perfectly appropriate touch, though I am not sure Morton and colleagues would see it that way.

Finally, I realize a lot of what I raise above is critical, but I want to end on a positive note to circle back to my opening comments. I think this manuscript and study could make a major contribution to the literature on communication and what factors influence vocal complexity in species. The authors did a huge amount of work here, tackling an important question with sophisticated approaches to vocal complexity. The manuscript and study are already quite strong

– I hope my comments, questions, and concerns are helpful in making the manuscript even stronger.

Referee: 2

Comments to the Author(s)
I like the comparative approach in this study. I think it is very useful to compare the use of vocalizations across closely related species in this way. The data sets are impressive and the statistical analyses seem sophisticated and sound. The differences are interesting and worthy of discussion. My main concern is with the overall presentation. It seems to try a bit too hard to present things as fitting with the social complexity hypothesis for communicative complexity. I would prefer to see the results presented in a bit more exploratory way and the fit with the complexity hypothesis discussed as one possibility. These are interesting results but I am not sure we can say so definitively what they mean.

Here are a few reasons why I don't see the data fitting with the hypothesis as cleanly as is presented. First, it is not clear what the diversity results really indicate. Looking at Table 4 shows that numbers of different call types used in the agonistic context are the same for Japenese+rhesus vs Tonkean+crested (only J+r use food calls and growls, only T+c use screech and soft grunt, the rest are shared). Also, the 3D figures indicate the J+r species have bigger clusters despite having fewer clusters. It seems that one could just as easily emphasize these features to indicate similarity in diversity. Also the affiliative calls show similar clustering differences as the agonistic clustering results (although slightly less clean) which doesn't support the prediction.

Furthermore, I don't quite agree with the use of graded calls as a measure of vocal flexibility and vocal complexity. I think the issue might lie with using potential information instead of realized information. With a coin flip, a balanced coin has more potential information than a biased coin because there is more uncertainty that can be resolved when the result is revealed. But it is the revealing of heads or tails that provides the information, not the ambiguity. The scoring of entropy used here is a measure of how difficult it is to categorize the calls, or the likelihood that calls fall acoustically between clusters. It is certainly possible that the presence of intermediate calls could indicate more potential information transfer but only if these calls communicated something different from what is communicated by the calls closer to the clusters. This would, essentially, indicate the presence of more call types. But that is not usually what happens with graded signals--typically the receiver uses contextual cues to place the intermediate calls in one of the clusters ("its heads"). The fact that the receiver has to do extra work to decode the signal seems consistent with the signal containing less, rather than more, information. In animal communication, information is usually used to refer to the reduction in uncertainty, not the uncertainty itself. An ambiguous signal reduces the uncertainty to a lesser extent. And intermediate signals (that have to be assigned by the receiver to one of the clusters) are ambiguous. I realized that these are difficult and contentious issues. I think it is certainly interesting that the degree of call gradation differs and this is worth discussing, I just think the issue could be discussed more carefully.

Referee: 3

Comments to the Author(s)
I enjoyed this paper and found it very clear to read. There was enough detail in the methods for it to be entirely replicable, from details about the study groups to equipment to analyses. My main question is around the handling of the affiliative context – it seems that the tolerant species also had more vocal flexibility and diversity in the affiliative context, but it's not really included in the discussion. It seems unsurprising that a species with more uncertainty in aggressive outcomes would also have more vocal complexity in affiliative contexts, as these are also contexts in which

the individuals are managing their social standing/relationships. Perhaps the authors could either explain why they didn't consider the results from the affiliative context to be compelling, or could put more detail into discussing those results.

Introduction: Overall, the introduction gives nice, clear descriptions that provides readers with the background needed to get the most out of this paper. The descriptions of macaque social behaviour and structures are well detailed, and supports the justification for comparisons within this genus. There were just two points where more detail would be useful: (1) the definition of structural factors, and (2) information theory.

Line 56: the definition of "structural factors" is not very clear. It will be difficult for readers who are less familiar with environmental vs structural factors.

Line 71: the following paragraphs do a good job of outlining the limitations of how we currently measure social and vocal complexity.

Line 91: should this be "a means" (grammatical question)? Also, please explain how information theory helps us to measure uncertainty. This statement makes it seem obvious, but if readers aren't familiar with information theory they won't follow this point.

Line 140: It took me a moment to parse how hypothesis 2 and 3 are different. I would suggest re-ordering Hypothesis 2 – " Greater similarity should occur in more closely related species, for any variable, and regardless of the social context, so we should find more differences between Tonkean and crested macaques on the one hand, and Japanese and rhesus macaques on the other, than within each of these species pairs across variables and contexts."

Methods:
Very clear descriptions of the facilities/habitats and group compositions! I have reviewed a couple of papers recently that were lacking this, so thanks for including! The methods are detailed to the degree that they are almost completely replicable. This is unusual and very welcome.

Line 202 : "Contexts were defined according to the behaviours that could occur in the 3 s before or after" It matters whether the context is before or after the call – can you clarify what you mean by "3s before or after"? What if one context happens in the 3s before and another context happens in the 3s after?

Line 270: "We can measure vocal diversity by the number of call types in the repertoire of a species" Unclear whether you are taking the repertoire from this reference or are using the method described in this paragraph. If the former, should report species' repertoire sizes here, and also would be good to quantify if the groups used all of their species' repertoire. If the latter, should clarify that you are doing this "as follows".

Discussion:

Line 389: In the results, you do report a significant difference in vocal diversity in the affiliative context as well – why are the differences that you found for the affiliative context not as important as the ones from the agonistic context? Also on line 412 you say that "No more patterns were uncovered regarding vocal diversity" – while on line 336 your results show that rhesus macaques had smaller number of clusters than the other species and that rhesus macaques had a smaller number of clusters than the Tonkean/crested pair. This seems to reflect the pattern for aggressive behaviour, although perhaps not as strongly.

It seems that for vocal diversity and vocal flexibility there is evidence that the tolerant species are higher than the non-tolerant species. I don't see why this is a problem for the social complexity hypothesis: a species that has less certainty may need more affiliative strategies as well as more

agonistic ones. On Line 414 you briefly posit this, and it seems quite unremarkable to say that affiliative interactions are more important than neutral ones. It feels like affiliative behaviours are being dismissed, because the focus was on aggressive behaviours, and that the role of affiliative behaviours in uncertain social systems is glossed over. Line 427 goes some way to expressing the importance of affiliative behaviours, so I wonder if it is more the way the discussion has been structured than the authors' actual interpretations?

Overall, I find this paper very compelling and I hope that my review helps the authors to bring out their key findings and interpretations. The final two paragraphs gave the paper a nice conclusion.

Author's Response to Decision Letter for (RSPB-2020-0439.R0)

See Appendix A.

Decision letter (RSPB-2020-0439.R1)

15-May-2020

Dear Dr Rebout

I am pleased to inform you that your manuscript entitled "Tolerant and intolerant macaques show different levels of structural complexity in their vocal communication" 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. Please note the typo spotted by the Board Member and correct on the proofs.

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

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

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

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

Paper charges

An e-mail request for payment of any related charges will be sent out after proof stage (within approximately 2-6 weeks). The preferred payment method is by credit card; however, other payment options are available

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

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

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

Associate Editor:
Board Member
Comments to Author:
Thank you very much for carefully and thoroughly responding to all of the reviewers' comments on your original submission. From my reading of your revised article I believe that you have addressed the reviewers' concerns and questions, and enhanced the clarity of your article's rationale and interpretation of your results.

Please note on line 252 I think I have spotted a typo. Specifically you state "Our samples resulted in 434 calls in 24 Japanese macaques (agonistic context: total number of calls, 79 & mean number of calls per female ± SD, 3.30 ± 377..." but I do not think the SD can be 377 given the total number of call is 79 for this sample. Surely a decimal point is missing here. Please correct this.

Appendix A

Dear Editor,

We would like to bring to your attention the revision of our manuscript entitled "Tolerant and intolerant macaques show different levels of structural complexity in their vocal communication".

We thank reviewers for taking the time to read our manuscript and comment on it. This has allowed us to significantly improve our manuscript. You will find below the detail of the revision.

Best regards,
Nancy Rebout

Reviewer(s)' Comments to Author:

Referee: 1

The authors compare four species of macaque in terms of individuals' levels of tolerance and vocal complexity in the species. The authors had a nice small comparative data set here in that one pair of macaques (Tonkean and crested) are tolerant whereas the other pair (rhesus and Japanese) are intolerant. The authors find that in general, the tolerant species exhibit greater complexity in major contexts of vocal signaling than the intolerant species.

This is a well-written manuscript and described a solid (though small in terms of number of species) comparative study. Some important strengths of the work include the breakdown of vocal signals into different contexts (affiliative, aggressive, neutral) and the assessment of vocal complexity in finer detail than most studies carry out. I do have a number of concerns about the work or the reporting of the work, however.

First, from the onset it was not clear why or how the phylogenetic hypothesis makes a different prediction regarding vocal complexity in comparison to the social complexity hypothesis, since it appears from the text that the two tolerant species are most closely related (of these four species) – see lines 131-137.
We have added the following sentence to be more explicit:
"Because of these phylogenetic distances, it can be expected that the vocal signals used by individuals will differ more between these two pairs of species than within each pair. However, such differences should apply indiscriminately to the various vocal variables and social contexts, contrary to the social complexity hypothesis which specifies that contrasts between species should depend on the variables and contexts." (l. 147-151).
The hypotheses and predictions tested in the study are then detailed in the last paragraph of the Introduction section.

Second, it was not clear to me why the authors would claim that the social complexity hypothesis would predict greater complexity in tolerant species for aggressive signals only, and not affiliative (I can see why differences would not be predicted for neutral signals). Individuals are still needing to manage the behavior of others through signaling in affiliative contexts, so it seems to me one would expect socially complex species to exhibit greater complexity in signals in those contexts as well as in aggressive / agonistic contexts, under the social complexity hypothesis. Gustison et al. (2012) found, for example, that geladas were similar to other and less socially complex species in terms of their vocal signals except for the

greater complexity of their affiliative and reconciliatory vocalizations. See:
• Gustison, M. L., A. le Roux, et al. (2012). "Derived vocalizations of geladas (Theropithecus gelada) and the evolution of vocal complexity in primates." Philosophical Transactions of the Royal Society B-Biological Sciences 367(1597): 1847-1859.
The authors seem later to admit as much in terms of the complexity of affiliative contexts, in lines 425-432.
Our belief was that agonistic interactions could have more important consequences that affiliative interactions. However, we admit that being skilful in cooperation is as much important as competition, as we wrote it in the Discussion section. Moreover, we agree with the reviewer that the affiliative context should involve more complex communication signals that neutral interactions that by definition do not involve social interactions. We have therefore modified several sentences of the Introduction section as follows:
"The need for complex communication signals is not necessarily the same in all social contexts [10]. In the agonistic context, animals need information to cope with the many potential outcomes of uncertain situations such as open contests between two or more individuals, which affects competition for resources and expose individuals to risk of injury. In the affiliative context, a wealth of communication signals can also help individuals to achieve the best solution from a variety of behavioural options and maintain their social relationships [25,35]. Significant interspecies differences in communication systems are to be expected in situations of competition and cooperation. On the contrary, no significant interspecies differences should occur in neutral circumstances – i.e. when individuals are not directly involved in a social interaction – that do not require the expression of a wide range of intentions." (l. 111-121)

Third, it is crucial to report the extent of recording work across dates and times for each species, rather than just in terms of sample duration when data were collected (see lines 180-187). I doubt this is the case in this study, but a worry would be that sampling effort across dates for rhesus and Japanese macaques, as opposed to Tonkean and crested macaques. If, for example, rhesus and Japanese were recorded in a relatively short time span, that could result in diminished contextual and motivational variation, contributing to decreased variation in signal structures.
The duration and dates of recording in each of the social groups are given in Table 1 of supplementary material S1. Sampling lasted 3-4 months in rhesus and Tonkean macaques, and 5-6 months in Japanese and crested macaques.

Fourth, the authors need to carry out inter-observer reliability measures to support the coding of behavior categories here. Hopefully they are able to assess reliability for coding of signals produced in the three different contexts (lines 205-210), though if that were done live, perhaps there is no way to assess reliability. Nonetheless, inter-observer reliability could easily be done for the measure of vocal diversity (lines 270-274).
The study of the four species was spread over 11 years, and the social contexts have been recognized live by the observers, so there is no way to assess the reliability of the records. However, the processing of vocal signals was insured by a single experimenter to ensure comparability of data across the four species. We now write, "The first author (N.R.) drew spectrograms using the software Raven Pro v1.4' (Cornell Lab of Ornithology, Center for Conservation Acoustics, Ithaca, NY, USA) with a 256 fast Fourier transform length and a Hanning window. With the same software, she measured the following variables, etc." (l. 231-234)

Additional comments by line number:

34. Perhaps "within" is better here than "between"? Or do the authors mean "among" instead?

We prefer to keep "between". This is explained in the Methods section: "We can measure signal uncertainty as the degree of gradation between call types [23]. We named *vocal flexibility* the degree of gradation between calls." (l. 307-308)

50. By "structural" do the authors mean the term in the way Hinde (1976, in journal Man) meant it?
No, Hinde (1976) used this term to mean "social structures". We ourselves use this term in a much broader sense. Structure refers to the organisation of the phenotype and its transformations during development and evolution. Most of the references in the paragraph refer to this meaning [3,4,5,6,8]. For the sake of clarity, we have completed the first sentence of the Introduction section, "When looking for the determinants of social evolution in animals, two main types of factors can be distinguished: external pressures coming from the environment and internal constraints arising from the structure of the phenotype". (l. 51-52)

77. Instead of "undermines" here, do the authors really mean "limits"?
Fixed. We now write "limits". (l. 79)

90. "The ability to produce uncertainty" seems an odd way of phrasing the idea for "Greater diversity among and within units of a system". The "to produce uncertainty" implies either an agency or a function that I am not sure the authors intend.
The two papers cited actually argue that the production of uncertainty is the main characteristic of complex systems. However, to avoid implying any agency, intention or function, we now write, "Uncertain outcomes appear to be the most important characteristic of complex systems [29,30]". (l. 92-93)

105-107. I don't think these two ideas perfectly align with one another, but wonder if a counter-argument to what the authors write here is that there would seem to be a great deal of honesty / reliability guarantees in signaling even in contest behavior – see the Bradbury & Vehrencamp chapters on this in their textbook or Searcy & Nowicki's book on the evolution of animal communication. If so, one might not necessarily expect much in the way of uncertainty in contest situations.
The question of whether communication signals are particularly reliable in conditions of open contest is a special issue, and we prefer to avoid making assumptions about this. We have modified this section in response to a previous comment by the reviewer. We now state that interactions occurring in agonistic and affiliative contexts involve a greater number of options and consequences than the neutral context where individuals are not directly involved in a social interaction (l. 111-121).

108-111. I understand the argument here for neutral situations, but don't understand it at all for affiliative situations. The authors should really develop this more as to why they hold this view, or consider the point above (and the Gustison 2012 reference) that affiliative situations, too, should result in greater uncertainty. In either case, a bit more discussion about the view regarding affiliative contexts would help a reader here.
As mentioned earlier, we now acknowledge that "In the affiliative context, a wealth of communication signals can also help individuals to achieve the best solution from a variety of behavioural options and maintain their social relationships [25,35]. Significant interspecies differences in communication systems are to be expected in situations of competition and cooperation. On the contrary, no significant interspecies differences should occur in neutral circumstances – i.e. when individuals are not directly involved in a social interaction – that do not require the expression of a wide range of intentions." (l. 114-121)
([25] refers to the article by Gustison et al., 2012)

143. Should be "Social Complexity hypothesis"
Fixed (l. 158).

144-147. The authors address this later in the discussion, but they communicate a directionality (social -> vocal) here that their data cannot address. It is equally likely in this study that vocal -> social and that vocally complex species (who evolved that complexity presumably for other reasons) are the only ones able to evolve complex social structures – this was the basic interpretation of the McComb & Semple comparative study the authors cite (although that study, too, could not determine directionality!).
We agree. We have replaced "should lead to" by "should be associated with" (l. 159) to avoid any assumption of directionality, in line with our comments in the Discussion section (l. 445-450).

252. Perhaps the authors could add a sentence as to why this cut-off of 0.7 is advocated in the study they cite? It seems rather arbitrary.
We have added the following sentence, "a simulation study showed that this is the value above which collinearity begins to bias model estimates, and is consequently the most commonly used threshold [58]". (l. 270-272)

262. I am not entirely sure what "stabilize cluster results" means – could the authors describe this in a sentence or two here?
By "stabilize cluster results", we mean "provide robust clustering". Indeed, PCA analyses reduce the dimensionality of the data set and provide more robust clustering. According to Ben-Hur and Guyon (2003), PCA improves the extraction of cluster structure with respect to stability. We now write, "We used Principal Component Analyses (PCA) to reduce the dimensionality of the data set and provide more stable clustering, which means that clustering outputs are less sensitive to outliers [61]." (l. 281-283)

395-407. I kept reading and rereading this section in comparison to the text in lines 390-394 and Fig 2, and still do not understand how the authors are making the interpretations they make here. From the results and Fig 2, it looks as though the differences seen between tolerant and intolerant species in terms of aggressive contexts are the same differences seen as for affiliative contexts, though perhaps not in as "statistically" clean a way as for aggressive contexts.
We agree. We have modified our comments in line with the other sections and the reviewer's comments:
"The diversity of call types was more extensive in Tonkean and crested macaques compared to Japanese and rhesus macaques in the context of aggression. We found a similar pattern in the affiliative context, although the difference between rhesus and Tonkean macaques was not statistically significant. On the other hand, we did not find similar contrasts between the two pairs of species in the neutral context. We also examined vocal flexibility by analysing the degree of gradation between groups of calls. We found the same type of demarcation between the Japanese/rhesus and the Tonkean/crested pairs in the agonistic and the affiliative contexts. As for vocal diversity, no difference appeared in the neutral context between both pairs of species." (410-418)

414-416. This sentence seems like a truism, rather than an "It could be . . ."
We have deleted this sentence.

421-425. I am on the "information" side of the "information / manipulation" debate, and so had a bit of a laugh when the authors followed a sentence referencing Morton with a sentence describing information in signals – a nice touch! And, for me, a perfectly appropriate touch,

though I am not sure Morton and colleagues would see it that way.
OK

Finally, I realize a lot of what I raise above is critical, but I want to end on a positive note to circle back to my opening comments. I think this manuscript and study could make a major contribution to the literature on communication and what factors influence vocal complexity in species. The authors did a huge amount of work here, tackling an important question with sophisticated approaches to vocal complexity. The manuscript and study are already quite strong – I hope my comments, questions, and concerns are helpful in making the manuscript even stronger.

Referee: 2

I like the comparative approach in this study. I think it is very useful to compare the use of vocalizations across closely related species in this way. The data sets are impressive and the statistical analyses seem sophisticated and sound. The differences are interesting and worthy of discussion. My main concern is with the overall presentation. It seems to try a bit too hard to present things as fitting with the social complexity hypothesis for communicative complexity. I would prefer to see the results presented in a bit more exploratory way and the fit with the complexity hypothesis discussed as one possibility. These are interesting results but I am not sure we can say so definitively what they mean.

Here are a few reasons why I don't see the data fitting with the hypothesis as cleanly as is presented. First, it is not clear what the diversity results really indicate. Looking at Table 4 shows that numbers of different call types used in the agonistic context are the same for Japenese+rhesus vs Tonkean+crested (only J+r use food calls and growls, only T+c use screech and soft grunt, the rest are shared). Also, the 3D figures indicate the J+r species have bigger clusters despite having fewer clusters. It seems that one could just as easily emphasize these features to indicate similarity in diversity. Also the affiliative calls show similar clustering differences as the agonistic clustering results (although slightly less clean) which doesn't support the prediction.
As written, Table 4 is for illustrative purposes only (l. 369), in the Supplementary material. Calls were assigned to call types by ear and/or visually, based on previous studies in macaques. However, according to authors, some call types may correspond to groups of several call types. For example, coo can be considered as a single category but in the Tonkean macaque three categories of coos have been reported (low, clear, and harsh coo: Masataka and Thierry 1993). The attribution of a call category is rather subjective. For this reason we have developed a method to quantitatively compare vocal complexity in different species. Thanks to this method, we were able to objectively classify calls, and this is where the demonstration lies. By presenting Table 4, we want to give readers an idea of the types of calls that were grouped by the clustering method, and so help them to relate our results to their prior knowledge.
The 3D figure in the Supplementary material indeed indicates that the Japanese and rhesus species have larger clusters although they have fewer clusters. However, the 3D graphs are a representation of a multi-dimensional space. As such, we do not recommend using the shape of the clusters as a meaningful and complete means of drawing conclusions about the clustering outputs. Although the cluster analysis takes all variations into account, the 3D graphs do not show all the variations. Again, these graphs are provided for illustrative purposes, and the demonstration should be sought in the results of the statistical tests.

Furthermore, I don't quite agree with the use of graded calls as a measure of vocal flexibility and vocal complexity. I think the issue might lie with using potential information instead of

realized information. With a coin flip, a balanced coin has more potential information than a biased coin because there is more uncertainty that can be resolved when the result is revealed. But it is the revealing of heads or tails that provides the information, not the ambiguity. The scoring of entropy used here is a measure of how difficult it is to categorize the calls, or the likelihood that calls fall acoustically between clusters. It is certainly possible that the presence of intermediate calls could indicate more potential information transfer but only if these calls communicated something different from what is communicated by the calls closer to the clusters. This would, essentially, indicate the presence of more call types. But that is not usually what happens with graded signals--typically the receiver uses contextual cues to place the intermediate calls in one of the clusters ("its heads"). The fact that the receiver has to do extra work to decode the signal seems consistent with the signal containing less, rather than more, information. In animal communication, information is usually used to refer to the reduction in uncertainty, not the uncertainty itself. An ambiguous signal reduces the uncertainty to a lesser extent. And intermediate signals (that have to be assigned by the receiver to one of the clusters) are ambiguous. I realized that these are difficult and contentious issues. I think it is certainly interesting that the degree of call gradation differs and this is worth discussing, I just think the issue could be discussed more carefully.

We believe that flexibility is a major component of complexity that has been neglected until now. We cite another paper [23] which has developed this view in detail: "From the simple statement that a system is a set of elements that are interrelated (von Bertalanffy, 1968), it comes that systems are composed of a variable number of elements (diversity), but also that the elements themselves can be variable (flexibility), and that they can associate in variable patterns (combinability)." This leads to consider several measures that can reflect three major dimensions of system complexity of systems: diversity, flexibility and combinability. The information theory of Shannon allows to quantify these different dimensions.

Regarding our measure of flexibility, we can see the degree of gradation in two ways. On the one hand, a higher degree of gradation leads to a higher number of call types mixing both call categories (i.e. vocal diversity) and subtypes/intermediates. On the other hand, a higher degree of gradation needs additional contextual information to interpret the signal, as intermediate calls may be ambiguous. In any case, this leads to a system with more options, more potential information, more uncertainty, and therefore with a higher degree of complexity.

The concept of 'information' is often misused because of its multiple meanings in everyday language. Information, in the sense of Shannon's information theory, refers to the degree of uncertainty that is synonymous with unpredictability. The underlying principle is that the greater the number of possible options, the greater the degree of uncertainty. In this theory, information is not synonymous with 'meaning', as it is often used in animal communication and in our opinion incorrectly, as it leads to the belief that information refers to the reduction of uncertainty. For clarity, we have added some explanatory sentences. We now write:

"Shannon's theory of information [31] provides a way to quantify diversity and flexibility in terms of uncertainty. This theory refers to what can be treated as a quantity of information which is here synonymous with a lack of *a priori* knowledge about the outcome of events, and therefore their unpredictability. More types of calls or more graded calls offer a greater number of options and, ultimately, the greater the number of options, the greater the uncertainty." (92-98)

Lastly, please note that we are not alone in advocating the use of flexibility in the assessment of vocal complexity. As we write it, "Some have proposed abandoning the idea of counting the number of calls to quantify vocal complexity, and instead using the degree of gradation of repertoires [24,28], i.e. flexibility in the acoustic structure of vocal signals. Since diversity and flexibility represent two different components of complexity, however, it seems that the best solution is to take both into account when characterising vocal complexity [23]." (l. 87-91)

Referee: 3

I enjoyed this paper and found it very clear to read. There was enough detail in the methods for it to be entirely replicable, from details about the study groups to equipment to analyses. My main question is around the handling of the affiliative context – it seems that the tolerant species also had more vocal flexibility and diversity in the affiliative context, but it's not really included in the discussion. It seems unsurprising that a species with more uncertainty in aggressive outcomes would also have more vocal complexity in affiliative contexts, as these are also contexts in which the individuals are managing their social standing/relationships. Perhaps the authors could either explain why they didn't consider the results from the affiliative context to be compelling, or could put more detail into discussing those results.

Introduction: Overall, the introduction gives nice, clear descriptions that provides readers with the background needed to get the most out of this paper. The descriptions of macaque social behaviour and structures are well detailed, and supports the justification for comparisons within this genus. There were just two points where more detail would be useful: (1) the definition of structural factors, and (2) information theory.

Line 56: the definition of "structural factors" is not very clear. It will be difficult for readers who are less familiar with environmental vs structural factors.
As mentioned above, we have completed the first sentence of the Introduction, "When looking for the determinants of social evolution in animals, two main types of factors can be distinguished: external pressures coming from the environment and internal constraints arising from the structure of the phenotype". (l. 51-52)

Line 71: the following paragraphs do a good job of outlining the limitations of how we currently measure social and vocal complexity.

Line 91: should this be "a means" (grammatical question)? Also, please explain how information theory helps us to measure uncertainty. This statement makes it seem obvious, but if readers aren't familiar with information theory they won't follow this point.
We have replaced "means" by "way" (l. 93)
We have also added some sentences explaining how information allows to measure uncertainty:
"Shannon's theory of information [31] provides a way to quantify diversity and flexibility in terms of uncertainty. This theory refers to what can be treated as a quantity of information which is here synonymous with a lack of *a priori* knowledge about the outcome of events, and therefore their unpredictability. More types of calls or more graded calls offer a greater number of options and, ultimately, the greater the number of options, the greater the uncertainty." (l. 92-98).

Line 140: It took me a moment to parse how hypothesis 2 and 3 are different. I would suggest re-ordering Hypothesis 2 – " Greater similarity should occur in more closely related species, for any variable, and regardless of the social context, so we should find more differences between Tonkean and crested macaques on the one hand, and Japanese and rhesus macaques on the other, than within each of these species pairs across variables and contexts."
Thank you, we have followed the suggestion (l. 154-158).

Methods:
Very clear descriptions of the facilities/habitats and group compositions! I have reviewed a couple of papers recently that were lacking this, so thanks for including! The methods are

detailed to the degree that they are almost completely replicable. This is unusual and very welcome.

Line 202 : "Contexts were defined according to the behaviours that could occur in the 3 s before or after" It matters whether the context is before or after the call – can you clarify what you mean by "3s before or after"? What if one context happens in the 3s before and another context happens in the 3s after?
We now specify, "Note that behaviour patterns could fluctuate before and after the emission of the call, but the context did not change." (l. 221-222)

Line 270: "We can measure vocal diversity by the number of call types in the repertoire of a species" Unclear whether you are taking the repertoire from this reference or are using the method described in this paragraph. If the former, should report species' repertoire sizes here, and also would be good to quantify if the groups used all of their species' repertoire. If the latter, should clarify that you are doing this "as follows".
This the latter option. We now specify it, "It is possible to measure *vocal diversity* by the number of call types in the repertoire of a species [12]. We ourselves measured it using the number of main categories of calls (i.e. groups of calls with similar acoustic characteristics) as follows." (l. 291-293)

Discussion:
Line 389: In the results, you do report a significant difference in vocal diversity in the affiliative context as well – why are the differences that you found for the affiliative context not as important as the ones from the agonistic context? Also on line 412 you say that "No more patterns were uncovered regarding vocal diversity" – while on line 336 your results show that rhesus macaques had smaller number of clusters than the other species and that rhesus macaques had a smaller number of clusters than the Tonkean/crested pair. This seems to reflect the pattern for aggressive behaviour, although perhaps not as strongly.
It seems that for vocal diversity and vocal flexibility there is evidence that the tolerant species are higher than the non-tolerant species. I don't see why this is a problem for the social complexity hypothesis: a species that has less certainty may need more affiliative strategies as well as more agonistic ones. On Line 414 you briefly posit this, and it seems quite unremarkable to say that affiliative interactions are more important than neutral ones. It feels like affiliative behaviours are being dismissed, because the focus was on aggressive behaviours, and that the role of affiliative behaviours in uncertain social systems is glossed over. Line 427 goes some way to expressing the importance of affiliative behaviours, so I wonder if it is more the way the discussion has been structured than the authors' actual interpretations?
Reviewer 1 has made the same comments regarding our explanations and comments regarding these results, and we have therefore made the changes requested by both reviewers:
"The need for complex communication signals is not necessarily the same in all social contexts [10]. In the agonistic context, animals need information to cope with the many potential outcomes of uncertain situations such as open contests between two or more individuals, which affects competition for resources and expose individuals to risk of injury. In the affiliative context, a wealth of communication signals can also help individuals to achieve the best solution from a variety of behavioural options and maintain their social relationships [25,35]. Significant interspecies differences in communication systems are to be expected in situations of competition and cooperation. On the contrary, no significant interspecies differences should occur in neutral circumstances – i.e. when individuals are not directly involved in a social interaction – that do not require the expression of a wide range of intentions." (l. 111-121)
"The diversity of call types was more extensive in Tonkean and crested macaques compared to Japanese and rhesus macaques in the context of aggression. We found the same pattern in the

affiliative context, although the difference between rhesus and Tonkean macaques was not statistically significant. On the other hand, we did not find similar contrasts between the two pairs of species in the neutral context. We also examined vocal flexibility by analysing the degree of gradation between groups of calls. We found the same type of demarcation between the Japanese/rhesus and the Tonkean/crested pairs in the agonistic and the affiliative contexts. As for vocal diversity, no difference appeared in the neutral context between both pairs of species." (l. 410-418)

Overall, I find this paper very compelling and I hope that my review helps the authors to bring out their key findings and interpretations. The final two paragraphs gave the paper a nice conclusion.