Familiar and unfamiliar face recognition in crested macaques (*Macaca nigra*)

Jérôme Micheletta, Jamie Whitehouse, Lisa A. Parr, Paul Marshman, Antje Engelhardt and Bridget M. Waller

Article citation details
*R. Soc. open sci.* 2: 150109.
http://dx.doi.org/10.1098/rsos.150109

Review timeline
Original submission: 1 July 2014
1st revised submission: 10 March 2015
2nd revised submission: 17 April 2015
Final acceptance: 30 April 2015

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

Review History

RSOS-140150.R0 (Original submission)

Review form: Reviewer 1

Is the manuscript scientifically sound in its present form?
Yes

Are the interpretations and conclusions justified by the results?
Yes

Is the language acceptable?
Yes

Is it clear how to access all supporting data?
Yes. Raw data provided.

Do you have any ethical concerns with this paper?
No

Have you any concerns about statistical analyses in this paper?
No
Recommendation?
Accept with minor revision (please list in comments)

Comments to the Author(s)
The manuscript addresses the interesting topic of face processing in non-human primates and adds this knowledge by assessing the effect of familiarity on a MTS task in a previously untested macaque species. The paper is well written. The main weakness of this paper is the very small number of subjects and thus the difficulty in generalising the findings. However, the authors are careful not to draw any strong conclusions from the data and clearly point out the limits of the study in a thorough discussion. Given the sound methodology and the remit of the journal, I think the study appropriate for publication in Open Science.

L32. By bias you mean they were better at discriminating higher-ranking individuals?
L48. It is possible that not all primates rely on facial cues for individual recognition.
L122. Point out here that the order of task 1 and 2 was counterbalanced across subjects.

Results – why were a number of models not compared rather than just the global versus null model?

In the discussion, a p value of 0.089 is discussed as a trend towards increased accuracy in the UIR task but the trend of 0.064 towards taking longer to respond to a dominant exemplar in the UIR task is not discussed. The trend towards increased response time for dominant individuals in the UIR might reflect a hesitancy when viewing an unknown dominant individual and contradicts the statement that the results indicate that dominance states cannot be inferred from facial cues (L413). It would be more consistent to consider all trends as either non-significant results or possible small effects.

Review form: Reviewer 2

Is the manuscript scientifically sound in its present form?
No

Are the interpretations and conclusions justified by the results?
No

Is the language acceptable?
Yes

Is it clear how to access all supporting data?
The paper does not make mention to the supplementary data file.

Do you have any ethical concerns with this paper?
No

Have you any concerns about statistical analyses in this paper?
Yes

Recommendation?
Reject
Comments to the Author(s)
The authors examined face recognition in crested macaques. They used a computerized match to sample paradigm and presented both familiar and unfamiliar conspecific faces as stimuli in the task. There were no differences in performance matching familiar or unfamiliar faces. Though performance in general was not terribly high.

One strength of the paper is that it does add to the body of knowledge concerning face discrimination and recognition in non-human species by testing a novel species. The field has been fairly limited in that regard. However, there are some serious issues with the way in which the study was designed and the conclusions the authors would like to make based on the design (addressed at the end).

Methods
Parts of the methods section would benefit from clarification about specific details of the design and methodology. For instance, it was not clear in the text the number of test trials versus the number of control trials presented in each session. When reading “24 unique trials repeated twice” in the text, it was assumed that this meant 24 test trials, when in fact according to Fig 3 this may actually have been 40 control trials and 8 test trials. The figure definitely helps clarify this, but the text could use modification to make this clearer as well.

Lines 178-179 – “after the initial response, two comparison...” What is this “initial response”? Were subjects required to select the sample image? This is not included in the description of the procedure.

Procedure – it was not clear how many photos were used to depict each individual, in either the unfamiliar or familiar stimulus sets. This information should be provided.

Training – virtually no information is provided about the training process, namely how long it took to train subjects. It is also not clear whether any of the images used in the training process were also used during testing. Or if the individuals depicted in the training process were also used during testing, although with new photos.

Did the control trials use the same images used during test trials?

Line 246 – perhaps change the word “but” in this sentence to “expect”, if it is to mean that the females did not see images of themselves.

Data analysis – reaction time, lines 302-303 - After removal of outliers, visual inspection was used to determine whether reaction times appeared normally distributed. This should be tested statistically. Reaction times are often highly skewed and require transformation based on the degree of skewness. Therefore, little can be concluded from any analyses performed using raw reaction time values if this was not statistically confirmed to be normally distributed.

Results, lines 378-380. The authors say that the subjects performed slightly better when they shared a strong bond with the match rather than a weak bond. While they do acknowledge that this was not statistically significant, it is actually nowhere near significant with a p-value of 0.434. I assume their statement comes from a visual inspection of figure 7. However, this statement (and interpretation) is highly misleading and not representative of the actual data.

Results, lines 393-394. The sentence is a bit confusing – “Subjects tended to take longer to complete a trial if the sample and match individuals were higher-ranking than the foil individual”. The sample and the match are the same individual, yes? Then perhaps removing the ‘s’ from the word ‘individuals’ would make that clearer.

At a couple of points in the results you refer to Table X – is this in the supplementary materials?
Discussion – the first line of the discussion states that it was demonstrated that crested macaques can discriminate faces of familiar and unfamiliar conspecifics. However, I am not sure where in the results they ever clearly demonstrated this. Rarely did subjects reach the authors’ arbitrarily set level of 75% correct. Performance appeared to be more consistently around 60-65%, which is just above chance levels and not entirely convincing that subjects could match familiar and unfamiliar conspecific faces.

That said, this likely is due to the testing procedure that was implemented by the authors. The current design uses a small, limited number of trials and exemplars to examine whether subjects were able to correctly match the identity of the individuals depicted in the images, and then repeatedly presents the same trials over and over again. Other research (e.g. Katz et al 2002; Wright & Katz 2006; Truppa et al 2010) has shown that when you present limited exemplars, such as what was done here, this reduces the likelihood of subjects learning the concept and generalizing to novel exemplars. While subjects here may know the match-to-sample rule and were able to use it to match identical images of faces or clip art, it is highly unlikely they understood they were to match using the concept of identity. And therefore did not generalize identity of the individual in the sample across multiple viewpoints to reliability or consistently select the correct match image. This is reflected in less than remarkably positive performance, highly variable performance, both within and between subjects, and that performance declined over time. Therefore, one cannot really conclude much about actual face discrimination and recognition abilities in crested macaques from this study design.

Review form: Reviewer 3 (Helene Meunier)

Is the manuscript scientifically sound in its present form?
Yes

Are the interpretations and conclusions justified by the results?
Yes

Is the language acceptable?
Yes

Is it clear how to access all supporting data?
Yes it is.

Do you have any ethical concerns with this paper?
No

Have you any concerns about statistical analyses in this paper?
No

Recommendation?
Accept with minor revision (please list in comments)

Comments to the Author(s)
Review of MS RSOS-140150

General comment
This research paper focuses on face processing in a still understudied but very socially interesting species, the crested macaque. The authors investigate both if the crested macaques show better
performances for discriminating faces of familiar versus unfamiliar conspecifics and if social factors (i.e. dominance status and social bonds) could influence their accuracy in face recognition. Experimental protocols comply with ethical standard, subjects had unrestricted access to the testing unit and could come and go when they pleased. Statistical analysis responds to high technical standard. The main weakness of this study lies in the small number of subjects (3) but the statistical analyses take it into account and in spite of the little number of tested individuals, results are convincing. Conclusions are supported by the data and the authors are cautious in their conclusions in regards to the small number of subjects.

In summary, this research provides novel and interesting data and I thus recommend this manuscript for publication in Royal Society Open Science. I just have a few minor points listed below.

Introduction
The introduction is clear and well structured. However, you could mention that monkeys seem to have mental representations of the dominance hierarchy of their group members (e.g. Bovet & Washburn 2003; Cheney & Seyfarth 1990, 2007).

Methods
The methodology is globally well described and detailed.

L177: please specify how you trained the subjects to touch the screen three times in rapid succession.

L188: please specify (if true, but I suppose/hope it is) that the pictures of unknown conspecifics you used in the training procedure was pictures of different individuals than the ones you used in the UIR task.

L202: replace “of the five individuals subject to the experiment.” by “of five individuals, including our three subjects and the two other group members.”

L226: not clear, please specify here that the 24 unique trials include both tests trials and controls for more legibility. (idem for line 253).

Figure 3: In the phase 2 of Dru & Sat for FIR tests, you wrote “12 test trials + 26 controls”. Did you mean “12 test trials + 36 controls”?

L266: replace “of the subject” by “of the subjects”.

Results
L347: what do you mean by “visual inspection”? Please explain. I would really appreciate to have access to Figure S1.

Discussion
L458: please insert “our” in front of “crested macaques”.
Hélène MEUNIER
PhD in Ethology-Primatology
Research and Welfare Manager
Primatology Centre of Strasbourg University
Fort Foch
67207 Niederhausbergen
France
+33 (0) 3 88 13 78 78
hmeunier@unistra.fr
Decision letter (RSOS-140150)

18-Nov-2014

Dear Dr Micheletta:

Manuscript ID RSOS-140150 entitled "Familiar and unfamiliar face recognition in crested macaques (Macaca nigra)" which you submitted to Royal Society Open Science, has been reviewed. The comments from reviewers are included at the bottom of this letter.

In view of the criticisms of the reviewers, the manuscript has been rejected in its current form. However, a new manuscript may be submitted which takes into consideration these comments.

Please note that resubmitting your manuscript does not guarantee eventual acceptance, and that your resubmission will be subject to peer review before a decision is made.

You will be unable to make your revisions on the originally submitted version of your manuscript. Instead, revise your manuscript and upload the files via your author centre.

Once you have revised your manuscript, go to https://mc.manuscriptcentral.com/rsos and login to your Author Center. Click on "Manuscripts with Decisions," and then click on "Create a Resubmission" located next to the manuscript number. Then, follow the steps for resubmitting your manuscript.

Your resubmitted manuscript should be submitted by 18-May-2015. If you are unable to submit by this date please contact the Editorial Office.

I look forward to a resubmission.

Sincerely,
Emilie Aime
Senior Publishing Editor, Royal Society Open Science
openscience@royalsociety.org

Author's Response to Decision Letter for (RSOS-140150)

See Appendix A.

RSOS-150109.R1 (Revision)

Review form: Reviewer 3 (Helene Meunier)

Is the manuscript scientifically sound in its present form? Yes

Are the interpretations and conclusions justified by the results? Yes
Is the language acceptable?
Yes

Is it clear how to access all supporting data?
Yes

Do you have any ethical concerns with this paper?
No

Have you any concerns about statistical analyses in this paper?
No

Recommendation?
Accept as is

Comments to the Author(s)
Authors took all my comments into account and corrected their manuscript according to my suggestions, so I recommend this paper for publication in Royal Society Open Science.

Hélène MEUNIER
Primatology Centre of Strasbourg University
Fort Foch
67207 Niederhausbergen
France
+33 (0) 3 88 13 78 78
hmeunier@unistra.fr

Review form: Reviewer 2

Is the manuscript scientifically sound in its present form?
Yes

Are the interpretations and conclusions justified by the results?
Yes

Is the language acceptable?
Yes

Is it clear how to access all supporting data?
Yes

Do you have any ethical concerns with this paper?
No

Have you any concerns about statistical analyses in this paper?
No

Recommendation?
Accept with minor revision (please list in comments)

Comments to the Author(s)
The authors have addressed the main concerns with this revision.
I do have a couple other comments, though, that would help strengthen/clarify the points of the paper.

For Figures 7 and 8, please specify in the captions that only familiar (FIR) trials are depicted.

I am hesitant about Figure 4 because if one just looks at that figure, they would conclude that participants did not perform well on this task, especially for FIR trials, as it shows mean performance of 50% over all trials. This is somewhat counter to the fact that you explicitly state that subjects performed above chance on all tasks. Therefore, what one concludes from the figure may not be what you want the take home message of the study to be. It may even be beneficial to remove this figure entirely.

Lastly, the authors analyzed whether there was an effect of dominance in the UIR trials. How was dominance determined? Please provide information about this in the methods, such as where you specify how the dominance and social bonds of the participant group was calculated (lines 145-151).

Decision letter (RSOS-150109)

15-Apr-2015

Dear Dr Micheletta

On behalf of the Editor, I am pleased to inform you that your Manuscript RSOS-150109 entitled "Familiar and unfamiliar face recognition in crested macaques (Macaca nigra)" has been accepted for publication in Royal Society Open Science subject to minor revision in accordance with the referee suggestions. Please find the referees' comments at the end of this email.

The reviewers and Subject Editor have recommended publication, but also suggest some minor revisions to your manuscript. Therefore, I invite you to respond to the comments and revise your manuscript.

- Ethics statement
  If your study uses humans or animals please include details of the ethical approval received, including the name of the committee that granted approval. For human studies please also detail whether informed consent was obtained. For field studies on animals please include details of all permissions, licences and/or approvals granted to carry out the fieldwork.

- Data accessibility
  It is a condition of publication that all supporting data are made available either as supplementary information or preferably in a suitable permanent repository. The data accessibility section should state where the article's supporting data can be accessed. This section should also include details, where possible of where to access other relevant research materials such as statistical tools, protocols, software etc can be accessed. If the data has been deposited in an external repository this section should list the database, accession number and link to the DOI for all data from the article that has been made publicly available. Data sets that have been deposited in an external repository and have a DOI should also be appropriately cited in the manuscript and included in the reference list.

- Competing interests
  Please declare any financial or non-financial competing interests, or state that you have no competing interests.
• Authors’ contributions
All submissions, other than those with a single author, must include an Authors’ Contributions section which individually lists the specific contribution of each author. The list of Authors should meet all of the following criteria; 1) substantial contributions to conception and design, or acquisition of data, or analysis and interpretation of data; 2) drafting the article or revising it critically for important intellectual content; and 3) final approval of the version to be published.

All contributors who do not meet all of these criteria should be included in the acknowledgements.

We suggest the following format:
AB carried out the molecular lab work, participated in data analysis, carried out sequence alignments, participated in the design of the study and drafted the manuscript; CD carried out the statistical analyses; EF collected field data; GH conceived of the study, designed the study, coordinated the study and helped draft the manuscript. All authors gave final approval for publication.

• Acknowledgements
Please acknowledge anyone who contributed to the study but did not meet the authorship criteria.

• Funding statement
Please list the source of funding for each author.

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 (i.e. by the 24-Apr-2015). If you do not think you will be able to meet this date please let me know immediately.

To revise your manuscript, log into https://mc.manuscriptcentral.com/rsos and enter your Author Centre, where you will find your manuscript title listed under "Manuscripts with Decisions". Under "Actions," click on "Create 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 referees and upload a file "Response to Referees" in "Section 6 - File Upload". You can use this to document any changes you make to the original manuscript. In order to expedite the processing of the revised manuscript, please be as specific as possible in your response to the referees.

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

1) A text file of the manuscript (tex, txt, rtf, docx or doc), references, tables (including captions) and figure captions. Do not upload a PDF as your "Main Document".
2) A separate electronic file of each figure (EPS or print-quality PDF preferred (either format should be produced directly from original creation package), or original software format)
3) Included a 100 word media summary of your paper when requested at submission. Please ensure you have entered correct contact details (email, institution and telephone) in your user account
4) Included the raw data to support the claims made in your paper. You can either include your data as electronic supplementary material or upload to a repository and include the relevant doi within your manuscript
5) Included your supplementary files in a format you are happy with (no line numbers, vancouver referencing, track changes removed etc) as these files will NOT be edited in production.
Once again, thank you for submitting your manuscript to Royal Society Open Science and I look forward to receiving your revision. If you have any questions at all, please do not hesitate to get in touch.

Best wishes

Emilie Aime
Senior Publishing Editor
openscience@royalsociety.org

Author's Response to Decision Letter for (RSOS-150109.R1)

See Appendix B.
Appendix A

Micheletta et al. Familiar and unfamiliar face recognition in crested macaques (Macaca nigra)

RSOS-140150

Familiar and unfamiliar face recognition in crested macaques (Macaca nigra)

Jérôme Micheletta, Jamie Whitehouse, Lisa A. Parr, Paul Marshman, Antje Engelhardt, Bridget M. Waller

We are grateful for the constructive comments of the editor and the two reviewers and will address them point-by-point. For clarity, the original comment is presented in italics and followed by our response (with corresponding line numbers in the new version of the manuscript when appropriate).

1. **Reviewer 1:**

   1.1. The manuscript addresses the interesting topic of face processing in non-human primates and adds this knowledge by assessing the effect of familiarity on a MTS task in a previously untested macaque species. The paper is well written. The main weakness of this paper is the very small number of subjects and thus the difficulty in generalising the findings. However, the authors are careful not to draw any strong conclusions from the data and clearly point out the limits of the study in a thorough discussion. Given the sound methodology and the remit of the journal, I think the study appropriate for publication in Open Science.

   Thank you!

   1.2. L32. By bias you mean they were better at discriminating higher-ranking individuals?

   Yes. We have reformulated this sentence to be more explicit: (l.33-35) “However, the subjects were better at discriminating higher-ranking familiar individuals, but not unfamiliar ones.”

   1.3. L48. It is possible that not all primates rely on facial cues for individual recognition.

   Yes. We did not mean to imply that vision is the only sense that can be used for individual recognition; only that individual recognition (regardless of the sensory modality used) is necessary to live in a complex group. We have reformulated this sentence to be clearer: (l.47-50) “In addition, primates often live in complex social groups with highly individualized and complex social relationships, and so need to accurately distinguish between individuals. Several species can use visual cues present on the face to discriminate individuals from still photographs”.

   1.4. L122. Point out here that the order of task 1 and 2 was counterbalanced across subjects.

   We have added this information as requested. (l.125-126) “The order of the tasks was counterbalanced across subjects”.

   1.5. Results – why were a number of models not compared rather than just the global versus null model?
Our aim was to test a set of predictors that were likely to influence the macaques’ performances given previous studies on the topic and with this species. We are not trying to find the best model, which is why we have not included a model selection analysis.

First, we compared the null model (no predictor) to the full model (including predictors) to show that the predictors included in the full model have a significant explanatory value [1,2]. This is necessary because predictors can be significant while the model is not actually different from the null model. Second, when categorical predictors with more than two levels are included in a model (e.g. stimulus sets), we need to compare the model including this predictor to a model without this predictor to assess the overall significance of the model. This is necessary because the default procedure implemented in the package we used (lme4 for R) only compares the values of the different categories to a reference category (e.g. stimulus set 1 vs. 2, 1 vs. 3, 1 vs. 4 etc.).

1.6. In the discussion, a p value of 0.089 is discussed as a trend towards increased accuracy in the UIR task but the trend of 0.064 towards taking longer to respond to a dominant exemplar in the UIR task is not discussed. The trend towards increased response time for dominant individuals in the UIR might reflect a hesitancy when viewing an unknown dominant individual and contradicts the statement that the results indicate that dominance states cannot be inferred from facial cues (L413). It would be more consistent to consider all trends as either non-significant results or possible small effects.

We have modified the discussion according to the reviewer’s comment. All p-values under 0.1 are discussed as trends. (l.420-423) “The macaques also tended to take longer to respond when viewing faces of unfamiliar high-ranking individuals, which suggests that they could infer some information regarding the social status of unfamiliar individuals from facial cues only”.

2. **Reviewer 2:**

2.1. The authors examined face recognition in crested macaques. They used a computerized match to sample paradigm and presented both familiar and unfamiliar conspecific faces as stimuli in the task. There were no differences in performance matching familiar or unfamiliar faces. Though performance in general was not terribly high.

We agree that performances were not terribly high. However, they are comparable to previous results obtained with other macaques tested in similar tasks [3,4]. As suggested by reviewer 2, it is possible that the small number of trials used in the study made it difficult for the macaques to attain higher performances. A consideration of the relative poor performance is now included in the discussion (l.488-506). We argue that the level of performance is likely to be the result of a motivation issue (see section 2.13 below for details), which is very hard to avoid in zoo settings.

2.2. One strength of the paper is that it does add to the body of knowledge concerning face discrimination and recognition in non-human species by testing a novel species. The field has been fairly limited in that regard. However, there are some serious issues with the way in which the study was designed and the conclusions the authors would like to make based on the design (addressed at the end).
We hope that the changes made throughout the manuscript, as well as our answers will be sufficient to address these concerns.

2.3. Parts of the methods section would benefit from clarification about specific details of the design and methodology. For instance, it was not clear in the text the number of test trials versus the number of control trials presented in each session. When reading “24 unique trials repeated twice” in the text, it was assumed that this meant 24 test trials, when in fact according to Fig 3 this may actually have been 40 control trials and 8 test trials. The figure definitely helps clarify this, but the text could use modification to make this clearer as well.

The text has been modified to clarify this point. (l.231 & l.260) “24 unique trials including test and control, repeated twice”. We have also added extensive details about the training to the supplemental materials (see ESM1).

2.4. Lines 178-179 – “after the initial response, two comparison…” What is this “initial response”? Were subjects required to select the sample image? This is not included in the description of the procedure.

The initial response consisted of touching the sample 3 times in rapid succession. This is clearly stated in the previous sentence (l.180-181): “First, we required subjects to orient towards a single image on the screen (hereafter, the sample) by touching it three times in rapid succession”.

2.5. Procedure – it was not clear how many photos were used to depict each individual, in either the unfamiliar or familiar stimulus sets. This information should be provided.

Each stimulus was made from a unique photo. There were as many unique photos as unique stimuli in each task (l.220-221 for the unfamiliar task): “We used twelve unique stimulus sets (36 unique pictures of 24 different individuals) in this task”. This information was indeed missing for the familiar task so we have now included it; Thank you for pointing this out (l.254-255): “We used 12 unique stimulus sets (36 unique pictures of 4 different individuals) in this task”.

2.6. Training – virtually no information is provided about the training process, namely how long it took to train subjects. It is also not clear whether any of the images used in the training process were also used during testing. Or if the individuals depicted in the training process were also used during testing, although with new photos.

We have added additional detailed information regarding the training process in the ESM1. The images used in the training we never used during testing and the individuals depicted in the training were never used in the tests. Specifically, we used pictures of crested macaques held in various UK zoo collections during training and pictures of wild individuals during the tests. This is now clearly stated in the manuscript (l.192-193).

2.7. Did the control trials use the same images used during test trials?

No, we used different images. This information is now included in the description of the procedure (Unfamiliar task: l.225-227; Familiar task: l.256-258): “In control trials, the
sample and match were identical photos of an unfamiliar conspecific (different to the ones used in the test trials) and the foil depicted an unfamiliar rhesus macaque”.

2.8. Line 246 – perhaps change the word “but” in this sentence to “expect”, if it is to mean that the females did not see images of themselves.

This has been corrected (l.252).

2.9. Data analysis – reaction time, lines 302-303 - After removal of outliers, visual inspection was used to determine whether reaction times appeared normally distributed. This should be tested statistically. Reaction times are often highly skewed and require transformation based on the degree of skewness. Therefore, little can be concluded from any analyses performed using raw reaction time values if this was not statistically confirmed to be normally distributed.

This was not explained clearly so we have now tried to improve it. The reaction times used as the dependent variable were log-transformed for the analysis (we did not analyse the raw reaction times). This is now clearly stated in the analysis section. (l.310) “Reaction times were log-transformed to approximate normal distribution.” Given the number of trial analysed, tests of normality before model construction are not usually recommended as the power of a Shapiro-Wilk (or Kolmogorov-Smirnov) test would be very high, and so even small deviations from normality will lead to a significant result.

The procedure we have followed is the common procedure recommended when using GLMMs. Following Zuur and colleagues [5] we assessed the validity of the model graphically (see model diagnostic plots below) by inspecting normality of the residuals of the DV, rather than assessing normality of the raw dependent variable. All visual inspections indicate that the model is valid.

2.10. Results, lines 378-380. The authors say that the subjects performed slightly better when they shared a strong bond with the match rather than a weak bond. While they do acknowledge that this was not statistically significant, it is actually nowhere near significant with a p-value of 0.434. I assume their statement comes from a visual
inspection of figure 7. However, this statement (and interpretation) is highly misleading and not representative of the actual data.

We have removed this statement from the results. (l.386-388) "In the FIR task, subjects’ performances were similar regardless of whether they shared a strong or weak bond with the match individual (β = 0.143, SE = 0.183, P =0.434, figure 7)."

2.11. Results, lines 393-394. The sentence is a bit confusing – “Subjects tended to take longer to complete a trial if the sample and match individuals were higher-ranking than the foil individual”. The sample and the match are the same individual, yes? Then perhaps removing the ‘s’ from the word ‘individuals’ would make that clearer.

This has been corrected.

2.12. At a couple of points in the results you refer to Table X – is this in the supplementary materials?

Despite repeated searches, we could not find any reference to Table X in the submitted version of the manuscript. We are only referring to table 1, which should be the only table in the manuscript. Table 1 is visible below the reference list (p29).

2.13. Discussion – the first line of the discussion states that it was demonstrated that crested macaques can discriminate faces of familiar and unfamiliar conspecifics. However, I am not sure where in the results they ever clearly demonstrated this. Rarely did subjects reach the authors’ arbitrarily set level of 75% correct. Performance appeared to be more consistently around 60-65%, which is just above chance levels and not entirely convincing that subjects could match familiar and unfamiliar conspecific faces. That said, this likely is due to the testing procedure that was implemented by the authors. The current design uses a small, limited number of trials and exemplars to examine whether subjects were able to correctly match the identity of the individuals depicted in the images, and then repeatedly presents the same trials over and over again. Other research (e.g. Katz et al 2002; Wright & Katz 2006; Truppa et al 2010) has shown that when you present limited exemplars, such as what was done here, this reduces the likelihood of subjects learning the concept and generalizing to novel exemplars. While subjects here may know the match-to-sample rule and were able to use it to match identical images of faces or clip art, it is highly unlikely they understood they were to match using the concept of identity. And therefore did not generalize identity of the individual in the sample across multiple viewpoints to reliability or consistently select the correct match image. This is reflected in less than remarkably positive performance, highly variable performance, both within and between subjects, and that performance declined over time. Therefore, one cannot really conclude much about actual face discrimination and recognition abilities in crested macaques from this study design.

We acknowledge that the size of our stimulus set is small compared to the studies mentioned here, which could be problematic, and have included this in the discussion (l.488-506, see also section 2.1). We would argue, however, that it is unlikely that this means the macaques did not match the pictures based on identity. The alternative, lower-level explanation would be that they learnt the visual appearance of the stimulus sets and used associative learning to complete the task. If this was the case, then the percentage of
success should be a lot higher since this is a basic skill and the number of association to remember is not high. If the macaques did not understand the task at all and were only responding randomly then their performances would be centred around 50%. Although not terribly high, the performances of the macaques tested in this study were consistently above chance level of 50%.

We believe that the explanation for the macaques’ performances might lie in the specificities of our experimental setting and the general procedures we used. Specifically, we tested animals who belonged to a social group and whose group-mates were often waiting near the testing area (outdoor platform, see figure 1). These social distractions were unavoidable in this study and may impact negatively on the macaques’ performances. It is also worth mentioning that unlike working in research facilities, working in zoos requires a number of compromises. In our case, for dietary reasons, we were limited in the number of trials we could conduct each days. We were also limited in the types of food items that we could use as rewards. All these reasons (social distractions, limited number of trials, limited food rewards) could have reduced the macaques’ motivations and therefore lowered their performances.

It is clear that these issues can be problematic when studying animals' cognition. However, if we want to further our knowledge by studying novel species, we need to be ready to relax some constraints in the design of our experiments even if it comes at a cost in terms of reduced controlled conditions. Only then will researchers be able to access and test the abilities of rare, endangered and socially interesting species (such as crested macaques), which are often housed in zoos under various constraints. We believe that access to these populations is vital to address important questions regarding the evolution and function of socio-cognitive skills.

3. **Reviewer 3:**

   3.1. *This research paper focuses on face processing in a still understudied but very socially interesting species, the crested macaque. The authors investigate both if the crested macaques show better performances for discriminating faces of familiar versus unfamiliar conspecifics and if social factors (i.e. dominance status and social bonds) could influence their accuracy in face recognition. Experimental protocols comply with ethical standard, subjects had unrestricted access to the testing unit and could come and go when they pleased. Statistical analysis responds to high technical standard. The main weakness of this study lies in the small number of subjects (3) but the statistical analyses take it into account and in spite of the little number of tested individuals, results are convincing. Conclusions are supported by the data and the authors are cautious in their conclusions in regards to the small number of subjects. In summary, this research provides novel and interesting data and I thus recommend this manuscript for publication in Royal Society Open Science. I just have a few minor points listed below.*

   Thank you!

   3.2. *The introduction is clear and well structured. However, you could mention that monkeys seem to have mental representations of the dominance hierarchy of their group members (e.g. Bovet & Washburn 2003; Cheney & Seyfarth 1990, 2007).*
We have followed this recommendation: (l.90-92) “a number of playback studies showed some species seem to have mental representations of kin and dominance relationships within their group [16,23].”

3.3. L177: please specify how you trained the subjects to touch the screen three times in rapid succession.

Detailed information regarding the training of the animals is now included in the ESM. One of the very first steps of the training was to train the macaques to touch the screen once to obtain a food reward. Once they mastered this simple task, we required the macaques to touch the screen three times within 5 s to progress. They all did this spontaneously since they were used to get a food item after pressing the screen once. When no reward was obtained, they just kept touching the screen. After this, the next stages of the MTS procedure were added incrementally. The initial training was conducted with pictures of the actual food rewards, the later stages with cliparts, the final stages with pictures of conspecifics.

3.4. L188: please specify (if true, but I suppose/hope it is) that the pictures of unknown conspecifics you used in the training procedure was pictures of different individuals than the ones you used in the UIR task.

Yes, the pictures used in the training phase were different from the pictures used in the test phases. Specifically, we used pictures of crested macaques held in various UK zoo collections during training and pictures of wild individuals during the tests.

3.5. L202: replace “of the five individuals subject to the experiment.” by “of five individuals, including our three subjects and the two other group members.”

This has been corrected (l.206-207).

3.6. L226: not clear, please specify here that the 24 unique trials include both tests trials and controls for more legibility. (idem for line 253).

This has been corrected (l.231 and 260).

3.7. Figure 3: In the phase 2 of Dru & Sat for FIR tests, you wrote “12 test trials + 26 controls”. Did you mean “12 test trials + 36 controls”?

Yes, these typos have been corrected.

3.8. L266: replace “of the subject” by “of the subjects”.

This has been corrected (l.274).

3.9. L347: what do you mean by “visual inspection”? Please explain. I would really appreciate to have access to Figure S1.

By visual inspection, we mean that by looking at the pictures side by side (as in figure S1, see also below) we could not see any obvious difference (e.g. brightness, contrast, sharpness, head orientation etc.) between the stimulus sets on which macaques were
highly/poorly successful and the others. Please note that using Photoshop's colour match tool, we matched the pictures for contrast, brightness and that the same camera was used to take all pictures. Sharpness was also adjusted with Photoshop.

We apologise if the electronic supplementary material was not accessible. I am not sure why this was the case. I can confirm that the ESM has been successfully uploaded in the revised version. For ease of access, the figure is also included at the end of this document.

3.10. L458: please insert “our” in front of “crested macaques”.

We have followed the reviewer’s suggestion but to avoid ambiguity regarding ownership of the macaques, we have reformulated the reviewer’s suggestion as: (l.466) “the crested macaques tested in this study”.
The macaques performed poorly on the stimulus sets framed in red and were highly successful on the stimulus sets framed in green. See also figure 5.
Familiar and unfamiliar face recognition in crested macaques (*Macaca nigra*)

Jérôme Micheletta, Jamie Whitehouse, Lisa A. Parr, Paul Marshman, Antje Engelhardt, Bridget M. Waller

We are grateful for the constructive comments of the editor and the two reviewers and will address them point-by-point. For clarity, the original comment is presented in italics and followed by our response (with corresponding line numbers in the new version of the manuscript when appropriate).

1. **Reviewer 3:**
   1.1. *Authors took all my comments into account and corrected their manuscript according to my suggestions, so I recommend this paper for publication in Royal Society Open Science.*

   Thank you!

2. **Reviewer 2:**
   2.1. *The authors have addressed the main concerns with this revision. I do have a couple other comments, though, that would help strengthen/clarify the points of the paper.*

   Thank you, we have addressed your comments below.

   2.2. *For Figures 7 and 8, please specify in the captions that only familiar (FIR) trials are depicted.*

   This has been added to the captions of figure 7 and 8.

   2.3. *I am hesitant about Figure 4 because if one just looks at that figure, they would conclude that participants did not perform well on this task, especially for FIR trials, as it shows mean performance of 50% over all trials. This is somewhat counter to the fact that you explicitly state that subjects performed above chance on all tasks. Therefore, what one concludes from the figure may not be what you want the take home message of the study to be. It may even be beneficial to remove this figure entirely.*

   We would prefer to keep figure 4 and show the results as they are. Note that this figure is used to illustrate the comparison of the mean performances across all trials in both tasks (as indicated in the legend). This is different from performances of the last two consecutive sessions (the criterion to progress in the task) and which is given in table 1.

   2.4. *Lastly, the authors analyzed whether there was an effect of dominance in the UIR trials. How was dominance determined? Please provide information about this in the methods, such as where you specify how the dominance and social bonds of the participant group was calculated (lines 145-151).*
We have added a sentence to clarify this point. (l.150-151) “Similar measures of dominance were obtained from previous studies for the unfamiliar individuals used as stimuli [32,42].” Specifically, JM has collected data about the social relationships of the wild individuals used in this study for a previous project.