Peer Community In Ecology

Behavioral changes in the rapid geographic expansion of the great-tailed grackle

Esther Sebastián González *based on peer reviews by* ***Pizza Ka Yee Chow***, ***Francois-Xavier Dechaume-Moncharmont*** *and 1 anonymous reviewer*

Logan CJ, McCune KB, LeGrande-Rolls C, Marfori Z, Hubbard J, Lukas D (2023) Implementing a rapid geographic range expansion - the role of behavior changes. EcoEvoRxiv, ver. 3, peer-reviewed and recommended by Peer Community in Ecology. https://doi.org/10.32942/X2N30J

Submitted: 12 April 2023, Recommended: 28 August 2023

Cite this recommendation as:
Sebastián González, E. (2023) Behavioral changes in the rapid geographic expansion of the great-tailed grackle. *Peer Community in Ecology*, 100535. 10.24072/pci.ecology.100535

Published: 28 August 2023

While many species' populations are declining, primarily due to human-related impacts (McKnee et al., 2014), certain species have thrived by utilizing human-influenced environments, leading to their population expansion (Muñoz & Real, 2006). In this context, the capacity to adapt and modify behaviors in response to new surroundings is believed to play a crucial role in facilitating species' spread to novel areas (Duckworth & Badyaev, 2007). For example, an increase in innovative behaviors within recently established communities could aid in discovering previously untapped food resources, while a decrease in exploration might reduce the likelihood of encountering dangers in unfamiliar territories (e.g., Griffin et al., 2016). To investigate the contribution of these behaviors to rapid range expansions, it is essential to directly measure and compare behaviors in various populations of the species.

The study conducted by Logan et al. (2023) aims to comprehend the role of behavioral changes in the range expansion of great-tailed grackles (*Quiscalus mexicanus*). To achieve this, the researchers compared the prevalence of specific behaviors at both the expansion's edge and its middle. Great-tailed grackles were chosen as an excellent model due to their behavioral adaptability, rapid geographic expansion, and their association with human-modified environments. The authors carried out a series of experiments in captivity using wild-caught individuals, following a detailed protocol. The study successfully identified differences in two of the studied behavioral traits: persistence (individuals participated in a larger proportion of trials) and flexibility variance (a component of the species' behavioral flexibility, indicating a higher chance that at least some individuals in the population could be more flexible). Notably, individuals at the edge of the

population exhibited higher values of persistence and flexibility, suggesting that these behavioral traits might be contributing factors to the species' expansion. Overall, the study by Logan et al. (2023) is an excellent example of the importance of behavioral flexibility and other related behaviors in the process of species' range expansion and the significance of studying these behaviors across different populations to gain a better understanding of their role in the expansion process.

Finally, it is important to underline that this study is part of a pre-registration that received an In Principle Recommendation in PCI Ecology (Sebastián-González 2020) where objectives, methodology, and expected results were described in detail. The authors have identified any deviation from the original pre-registration and thoroughly explained the reasons for their deviations, which were very clear.

References:

Duckworth, R. A., & Badyaev, A. V. (2007). Coupling of dispersal and aggression facilitates the rapid range expansion of a passerine bird. Proceedings of the National Academy of Sciences, 104(38), 15017-15022. https://doi.org/10.1073/pnas.0706174104

Griffin, A.S., Guez, D., Federspiel, I., Diquelou, M., Lermite, F. (2016). Invading new environments: A mechanistic framework linking motor diversity and cognition to establishment success. Biological Invasions and Animal Behaviour, 26e46. https://doi.org/10.1017/CBO9781139939492.004

Logan, C. J., McCune, K., LeGrande-Rolls, C., Marfori, Z., Hubbard, J., Lukas, D. 2023. Implementing a rapid geographic range expansion - the role of behavior changes. EcoEvoRxiv, ver. 3 peer-reviewed and recommended by PCI Ecology. https://doi.org/10.32942/X2N30J

McKee, J. K., Sciulli, P. W., Fooce, C. D., & Waite, T. A. (2004). Forecasting global biodiversity threats associated with human population growth. Biological Conservation, 115(1), 161-164. https://doi.org/10.1016/S0006-3207(03)00099-5

Muñoz, A. R., & Real, R. (2006). Assessing the potential range expansion of the exotic monk parakeet in Spain. Diversity and Distributions, 12(6), 656-665. https://doi.org/10.1111/j.1472-4642.2006.00272.x

Sebastián González, E. (2020) The role of behavior and habitat availability on species geographic expansion. Peer Community in Ecology, 100062. https://doi.org/10.24072/pci.ecology.100062. Reviewers: Caroline Nieberding, Tim Parker, and Pizza Ka Yee Chow.

Reviews

Evaluation round #2

DOI or URL of the preprint: **https://doi.org/10.32942/X2N30J**
Version of the preprint: 2

Authors' reply, 31 July 2023

Dear Esther Sebastián González and Francois-Xavier Dechaume-Moncharmont,

We thank you very much for checking this version of the manuscript and we are so happy that you like our changes!

We responded to your remaining comments below and we revised the manuscript accordingly using track

changes. The track changes version of the manuscript was too large to upload at the PCI Ecology website, so here is the link to the .docx version: https://share.eva.mpg.de/index.php/s/MLLeMD5A89AxayA, and here is the link to the Google doc version: https://docs.google.com/document/d/1h_gRVffyakVIROQ4tnu3pSBNRfAW0rqjo4gNvnn1dAE/edit?usp=sharing.

We made sure to insert the change we promised to make in Response 6 in the last round (Results > Flexibility) into the current version. We also added a note at the top of "Post-study choices made since receiving in principle recommendation" that explains that this manuscript started out as a preregistration that was peer reviewed at PCI Ecology. We realized we hadn't made that explicit.

We wonder if it would be ok to add an additional analysis about the boat-tailed grackles in the Discussion? Since submitting the manuscript in the previous round, we were actually able to hire a video coder to code the boat-tailed grackle exploration videos. We ran an analysis, which shows that the boat-tailed grackles are less exploratory than the Tempe great-tailed grackles and similar to the Woodland great-tailed grackles (model outputs added to Table SM4 in Supplementary Material 4). We removed one sentence and added to two sentences in the Discussion to reflect this addition. If this is ok with you, then we will keep the text as it is. If not, we will revert the changes to the previous version.

We are really excited to share this piece of research with the world, and we thank you for all of your help along the way!

All our best,
Corina Logan (on behalf of all co-authors)

ROUND 2: Almost ready to recommendation

COMMENT 1: Dear Dr. Corina Logan and other authors,

Thank you very much for your detailed response letter and all the changes you have done in the text. I have carefully read the new version, together with your responses to the reviewers and I must say that you have done an excellent work! Two of your original reviewers were not available for this second round, but I am very happy with the way you have addressed their concerns and I do not see the need to try to find new reviewers.

Reviewer Francois-Xavier Dechaume-Moncharmont only has two very minor comments and, from my side, I can add other two very minor issues that I found while doing my final read:

- Page 2, lines 78-80. I think there is an extra "to" at the end of this sentence: Whereas, in exploration assays, the regular maintenance diet is provided far away from the novel element to assess the willingness to investigate novelty without the need to for food.
- Line 96: There word "that" is repeated twice

Thus, I am going to give you the chance to change these last details while I write my recommendation.

Looking forward to the last version of your study.

Regards,

Esther Sebastián González

RESPONSE 1: Thank you so much for going through all of the materials and assessing that we are almost ready for recommendation! We really appreciate your handling of this manuscript - thank you! Great catch on the extra "to" - we deleted it. We also deleted the extra "that" from line 96.

Reviews

Reviewed by Francois-Xavier Dechaume-Moncharmont, 12 Jul 2023 10:12

COMMENT 2: I carefully read through the new version of the MS and the response letter. The authors have

made appreciated efforts to convincingly address the comments made by the referees. The MS is now much clearer. Most issues have been convincingly addressed.

I am only surprised by response #14 "the most recent one we could find is from 2021". The CRAN developed function 'citation()' for that purpose (https://intro2r.com/citing-r.html). Using this function, the most recent citation reads R Core Team (2023). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Incidentally, I think the reviewers warrant acknowledgements, considering how they helped to improve the MS.

RESPONSE 2: Thank you so much for reading through the new version and the response! We appreciate your hard work on this manuscript, which has really improved it. Hmmm, it is odd, but when we run "citation()" in the R console, it still only gives us the 2021 version. Thank you for providing with the 2023 version - we updated our bibtex file so this is the one that now shows up in the references. We absolutely agree that reviewers should go in the Acknowledgements - we usually do this in the last version, so thank you for the reminder. We made the additions.

Decision by Esther Sebastián González, posted 27 July 2023, validated 27 July 2023

Almost ready to recommendation

Dear Dr. Corina Logan and other authors,

Thank you very much for your detailed response letter and all the changes you have done in the text. I have carefully read the new version, together with your responses to the reviewers and I must say that you have done an excellent work! Two of your original reviewers were not available for this second round, but I am very happy with the way you have addressed their concerns and I do not see the need to try to find new reviewers.

Reviewer Francois-Xavier Dechaume-Moncharmont only has two very minor comments and, from my side, I can add other two very minor issues that I found while doing my final read:

- Page 2, lines 78-80. I think there is an extra "to" at the end of this sentence: Whereas, in exploration assays, the regular maintenance diet is provided far away from the novel element to assess the willingness to investigate novelty without the need to for food.
- Line 96: There word "that" is repeated twice

Thus, I am going to give you the chance to change these last details while I write my recommendation.

Looking forward to the last version of your study.

Regards,

Esther Sebastián González

Reviewed by Francois-Xavier Dechaume-Moncharmont, 12 July 2023

I carefully read through the new version of the MS and the response letter. The authors have made appreciated efforts to convincingly address the comments made by the referees. The MS is now much clearer. Most issues have been convincingly addressed.

I am only surprised by response #14 "the most recent one we could find is from 2021". The CRAN developed function 'citation()' for that purpose (https://intro2r.com/citing-r.html). Using this function, the most recent citation reads R Core Team (2023). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Incidentally, I think the reviewers warrant acknowledgements, considering how they helped to improve the MS.

Evaluation round #1

DOI or URL of the preprint: https://doi.org/10.32942/X2N30J
Version of the preprint: 1

Authors' reply, 03 July 2023

Dear Esther Sebastián González, Francois-Xavier Dechaume-Moncharmont, Pizza Ka Yee Chow, and anonymous reviewer,

Thank you so much for taking the time to carefully review our post-study manuscript - we really appreciate the effort you invested in this! We revised the manuscript and responded to your comments (below) and we think it is much improved as a result.

In particular, we want to draw your attention to our Response 48. Since submitting this manuscript, we obtained the results from a different manuscript that involved the Arizona grackles, which showed that the flexibility manipulation the Arizona grackles underwent subsequently affected their behavior on the multiaccess box. Therefore, in the current manuscript, we now account for the potential that the performance of the manipulated birds might influence the cross population analyses. We did not account for this in the current preregistration because the Arizona flexibility manipulation was covered by a different preregistration at PCI Ecology. However, because the manipulation worked and affected other behaviors, we now changed our models to account for treatment by adding a term "treatment" (manipulated or not) to the analyses. We found that two of the four results changed: there is now no difference in exploration between populations and there is a difference in persistence with the edge population being more persistent. The flexibility result stayed the same: there is still no difference in average flexibility levels between populations. We also reran the power analyses to determine whether adding this term caused any changes in the minimum sample sizes. The minimum sample sizes either stayed the same or got smaller (because we got better at making our simulations more accurate) or, as in the case of exploration, the latencies by which the two sites must differ before a site difference can be detected increased.

In rerunning these analyses, we noticed that the multiaccess box data for the Arizona birds was not for the multiaccess log that we used with the California birds, but for a plastic multiaccess box that was only used in Arizona. We now corrected the error in the data so all multiaccess box data is from the log experiment. We apologize for this oversight! And we appreciate that the excellent feedback from you all helped us catch this. This changed the innovativeness result from previously showing more innovativeness in the edge population to now showing no population differences.

We also figured out how to conduct the Bayesian analyses to compare the variances between the sites, whereas previously we relied on visual assessments of the figures and an unregistered frequentist test in the case where it looked like there could be a difference in exploration variances (the test showed no difference). We conducted the Bayesian variance analyses and found that Woodland had a higher variance in phi (one of the two flexibility components). We updated the Abstract, Methods, Results, and Discussion sections accordingly.

One small note that we deleted the paragraph in the Discussion about how flexibility and innovativeness are causally linked in this species because it seemed a bit off topic.

Changes are shown in the tracked changes PDF at the PCI Ecology website and at the Rmd file at GitHub: https://github.com/corinalogan/grackles/blob/master/Files/Preregistrations/gxpopbehaviorhabitatq1.Rmd .

As a handy reference, here is the link to the preregistration that received in principle acceptance at PCI Ecology (https://doi.org/10.24072/pci.ecology.100062), and information on how to recommend and review Stage 2 manuscripts at PCI Registered Reports for recommenders (https://rr.peercommunityin.org/help/guide_for_recommenders#h_67596462364016136433909 05) and reviewers (https://rr.peercommunityin.org/help/guide_for_reviewers).

We thank you very much for your feedback on this article and we think the revised version is much improved due to your efforts!

All our best,
Corina (on behalf of all co-authors)

Round #1

by Esther Sebastián González, 30 May 2023 15:46
Manuscript: https://doi.org/10.32942/X2N30J version 1
COMMENT 1: Preprint merits a revision
Dear Dr. Corina Logan,
Thank you very much for your preprint. It has been reviewed by three experienced researchers and they all have underlined the importance of your work. However, they also see many points that need to be improved, especially to clarify terminology and methodology. The three reviewers have made an excellent job and have provided with detailed suggestions and discussion points that I think will largely improve the quality of your study. Because of the many issues raised, I cannot recommend your preprint in its present form, but I will be happy to read a revised version of it that takes into account or discusses all the points raised by the reviewers. I am looking forward to read this revised version.
Regards,
Esther
RESPONSE 1: Thank you very much for your assessment and for offering us the opportunity to revise and resubmit! Please find our detailed responses below.

Reviewed by Francois-Xavier Dechaume-Moncharmont, 29 May 2023 20:17
COMMENT 2: This article presents the results from a preregistered study on cognitive ecology in Grackle, from two populations in the USA, one being in the border of the expansion range of the species. Most notably, the objectives of the study were to assess the differences not only in mean but also in variance in cognitive performances between these two populations, in order to test the hypothesis of a link between invasibility and cognitive traits. The aim and scope of the MS were clear. The experimental design was appropriate and generally well presented. One of the salient features of this study is that it is based on a thorough a priori analysis of statistical power. The authors very rarely deviated from their pre-recorded protocols, and when they did, they stated so unambiguously. My overall impression about this manuscript is positive, and I recommend acceptation after revisions. Indeed, some points deserve clarification prior final acceptance.

RESPONSE 2: Thank you very much for your positive assessment! And thank you also for your detailed comments on how we can improve the manuscript, which we respond to below.

COMMENT 3: Major comments
Writing style. As a non-native English-speaking writer, I won't venture to criticize English or the syntax. Yet, the MS constantly alternates between past and present tense, sometimes in the same paragraph or even in

the same sentence ("we investigate… we aim for… we expected… Grackles are habituated… The doors were locked open…"). I strongly recommend rewriting the text in the past tense, so as to present the study and the results as a report on experiments that have already been carried out. Avoid the present tense, which should be confined to expressing general truths. The constant switching between present and past tense gives the very unpleasant impression of reading a MS hastily rewritten from the preregistered protocol.

RESPONSE 3: We were trying out writing the whole article in present tense as much as possible because our target journal is Peer Community Journal, which does not have tense requirements. However, either we failed at implementing this effectively or it is something that is too new to be widely accepted, therefore we changed the tense in the Methods, Results, and Supplementary Materials to past tense, as well as the results section of the Abstract. We kept the Introduction and Discussion in primarily present tense except when discussing our findings and those from other articles (per recommendations at https://www.nature.com/scitable/topicpage/effective-writing-13815989/).

COMMENT 4: The statistical analyses (methods and results) are particularly difficult to follow due to several issues. Some important analyses can only be understood when reading supplementary materials, and this information is not cited in the main text. For instance, the authors "estimated 89% compatibility intervals" (lines 358) without explanation. The reader must wait until supplementary materials (lines 604) to find a reference to Richard McElreath's book. The MS should be edited in order to present all the information necessary for the understanding or straightforward references to the corresponding section in supplementary materials.

RESPONSE 4: We apologize for the confusion. We added the McElreath citation to line 358 (and also throughout), and we went through the methods and results sections and added references to the supplementary materials that support these pieces. We have found with other post-study articles that reviewers prefer the bulk of the methods placed in supplementary material. However, if you think that some of the supplementary material would be better placed in the methods section, please let us know and we can move them accordingly. As it currently is, we tried to summarize the methods such that the reader could move through more quickly while having the option to read further in the supplementary material if they choose.

COMMENT 5: I also strongly recommend rewriting all the main text, tables and figures in order to avoid highly confounding mathematical notations. λ is used for both the deviation rate in flexibility analysis and a probability of approaching a novel environment in exploration analysis. ϕ is used for both the learning rate in flexibility analysis and the dispersion rate across bird in exploration analysis. $alpha$ in several model with different distribution law. The author acknowledges that confusion in the middle of the results section (lines 394-395). But that is clearly insufficient. The reader must be guided, and all possible elements of confusion must be eliminated, even is this thorough rewriting process is time consuming. Your time is as precious as the reader's. You write a paper because you are expecting to be read and understood (and possibly cited). A scientific article is not supposed to be a detective story for the reader. The acrimony of this comment reflects the difficulty I had in reading the MS.

RESPONSE 5: Thank you for pointing out this confusion - we are sorry that it caused so many difficulties for readers. We changed the lambda and phi in the Exploration analysis to L and P, respectively, to differentiate it from the flexibility components phi and lambda. We use alpha to refer to intercepts in all of the models and we think it is easier for readers to remember this if we keep the same symbol even across models even though the other within-model components differ. This is also the convention that is used in the Statistical Rethinking book that we learned these methods from McElreath (2020). We want to keep the notation similar to what is in Rethinking because it will make it much easier people who might want to modify/use our models in the future.

COMMENT 6: I acknowledge that the authors took great care in reporting effect sizes with compatibility intervals. This is unquestionably an excellent practice. Yet, these metrics alone are extremely hard to appreciate. For instance, the authors explain that "The flexibility measure is how many trials it takes them to reverse their color preference" (lines 160). This definition is straightforward and easy to understand from a biological perspective. But the flexibility is later only reported in terms of attraction score, learning rate (ϕ) and rate of deviating from learned attractions (λ). The estimated values of these metrics must certainly be reported in the MS, but the authors should not forget to report and illustrate biologically meaning metrics (so called "how many trials it takes them to reverse preference"). In addition, it is very important to explain in detail what this effect size indices mean, particularly when there are not trivial dependencies between them: "higher ϕ values meant lower λ values" (lines 234).

RESPONSE 6: Great point about placing the summary data per bird directly in the text. It was previously only available at the data repository so we now include a new table (Results > Table 2) where we present, per bird, the number of trials to reverse, their phi and lambda, the number of loci solved, etc. so readers can see what happened. Readers will be able to see how phi and lambda scores relate to the number of trials to reverse to get a sense of the scale, which should help with effect size interpretations. Additionally, the extra clarification we made to the Methods > Flexibility analyses should also make this clearer (see Response 76). After we had submitted the revised preprint to EcoEvoRxiv and it was approved, we realized that we could add some clarifying text to the Results section. We will add the clarifications below in the next revision:

Results > Flexibility: "With our sample size, we only have the power to reliably detect differences between the populations if they are larger than 0.01 for ϕ, which corresponds to a difference of 1% in how much individuals choose the rewarded option after they have just received a reward from this option. For λ, we would need a difference of at least 3, which corresponds to a 10% difference in how often an individual chooses the alternative option. The detection differences in ϕ and λ are based on our power analysis in Supplementary Material 2, summarized in Supplementary Material 1, and their correspondence with the number of trials to reverse comes from @blaisdell2021more."

COMMENT 7: I took time to careful read the MS and properly think about the methods, yet I still do not understand the analysis reported lines 374, about the innovation. The main issue come from the sentence: "the compatibility interval did not cross zero". As stated lines 246, the response value was to be found in the set \0, 1, 2, 3, 4\. Assuming random distribution of this measure, with most classical distributions, the difference between the average responses of the two populations, should not overlap zero. The null hypothesis should not be zero. But, here, I suspect that I misunderstood something. I recommend that the authors elaborate on the methodology. For instance, how did they compare the two populations? Did they use resampling (bootstrap) methods? If so, provide details about the procedure.

RESPONSE 7: We needed to answer the question about whether the two sites differed from each other in the means of the response variables. The most direct way to do this is to use an analysis called a contrast, which uses samples from the posterior distribution (McElreath 2020). We computed the estimated difference between the two sites (Tempe and Woodland) using contrasts, which gives the expected difference between the means (number of loci solved, exploration latency, etc.) of the two sites. This allows us to assess whether one site is systematically larger or smaller than the other by estimating what percentage of each sample of differences is either larger or smaller than zero. If 89% of the differences are larger than zero, then the older population has a larger mean, and if 89% of the differences are smaller than zero, then the edge population has a larger mean. We see now that this was explained in the Supplementary Material, but not very clearly in the Methods or Results. We now refer to the "contrast analysis" in the Results and we added the following to the Methods inside each analysis section (Flexibility, Innovation, Exploration, and Persistence): "We used an analysis called a contrast to assess whether one site was systematically larger or smaller than the other by

estimating what percentage of each sample of differences is either larger or smaller than zero. If 89% of the differences are larger than zero, then the older population has a larger mean, and if 89% of the differences are smaller than zero, then the edge population has a larger mean. If 89% of the differences cross zero, then we conclude that there is no strong difference between the sites."

COMMENT 8: Minor comments. Lines 80 and followings. "Our results DEMONSTRATE that the rapid geographic range expansion of great-tailed grackles is ASSOCIATED WITH individuals differentially expressing particular behaviors in the edge compared to the older population." Without speaking of the epistemological concerns about the fact that Science does not aim at demonstrating anything (such strong statements should be avoided), the study is purely correlative. The FUNCTIONAL link between cognitive skills and dispersion remains to be established. For instance, the difference could come from difference in ecological constraints in the two sites without any connection with the expansion history of the species. The study is sufficiently solid and new in itself, there is no need to overestimate it. I recommend that the authors reformulate the presentation of their results in a more neutral way.

RESPONSE 8: Thanks for catching this. We deleted this sentence because it was a statement including our results, which doesn't belong in the introduction.

COMMENT 9: Fig. 2. The far-right drawing (about exploration) is very difficult to understand. Please redraw and give additional explanations in the legend.

RESPONSE 9: We made a new figure based on a picture so it is similar to the others and we added the following description to the caption: "latency to approach a novel environment placed inside of the familiar environment and with regular food present, but not near the novel environment"

COMMENT 10: Lines 113 and following. In this paragraph, I would greatly appreciate finding information about size, mass and body conditions (Peig & Green 2009) of the birds from the two populations. Difference in body conditions would be particularly important when discussing differences in ecological success in between the two populations.

RESPONSE 10: We, too, have questions about how physical differences (as well as hormones and immunity) vary between the populations. However, these questions are planned for different articles farther in the future. The current article received in principle acceptance and is based on the behavioral differences we are investigating and, therefore, we will stay with the original, pre-approved plan.

COMMENT 11: Lines 150. The bullet points are useless here. They can simply be presented as plain sentences in regular paragraphs.

RESPONSE 11: We removed the bullets.

COMMENT 12: Lines 150. Avoid hypertext link in a MS. Beside, the protocol is crucial and must be included in the article, possibly as supplementary material, and not as an unreliable googledoc…

RESPONSE 12: We moved the protocols from the google doc to the new Supplementary Material 5 section.

COMMENT 13: Throughout the MS, the reference "Logan et al. 2023" is ambiguous since there are two possible papers corresponding to this abbreviation in the reference section. The main text should be thoroughly edited.

RESPONSE 13: Thanks for catching this - it was coming from a setting when we were exporting from the Rmd file to the PDF file. It was distinguishing between the two references using first name initials, so one was "Logan et al. (2023)" and the other was "C. Logan et al. (2023)". We were able to change this so it lists them as a and b by changing the style of the references.

COMMENT 14: Lines 197 and following. The authors provide a long inventory of the R packages used (psych, irr, rethinking, rstan, knitr, dplyr, tidyr, cmdstanr, DHARMa, lme4, Rcpp), without the reader knowing for what use and for what precise analysis. Be more specific and give sufficient information about the most important packages. Some accessory packages are clearly less critical (knitr, dplyr, tidyr) because they are mainly used to format data or code. They can be omitted from the main text and only quoted in the source code if they are relevant. In addition, I am surprised by the largely outdated reference (2017) for R language while the cited version of the software (4.1.2) was published in November 2021.

RESPONSE 14: We added descriptions of what the various packages were used for and we cited all packages used to ensure proper credit was given to these generous open source developers. We updated the R citation - the most recent one we could find is from 2021.

COMMENT 15: Line 239, Table 1, Fig 3, Table 4, etc. Use consistent notation (either the Greek letter or its transcription in Latin alphabet, e.g. phi, lambda) throughout the MS.

RESPONSE 15: Thank you for pointing out this inconsistency. We now changed to consistently using the Greek letters except for Tables 2, 3, and SM1 where we were not able to get the csv file to interpret these symbols.

COMMENT 16: Lines 242. This information about R version is a repetition of information already presented on the previous page.

RESPONSE 16: We removed the R information here, thanks for catching this.

COMMENT 17: Lines 316. Remove useless brackets.

RESPONSE 17: We fixed the formatting, thank you for pointing this out.

COMMENT 18: Table 2. Since the analyse of the study was to compare the value and their compatibility interval, I recommend to present all ϕ values in consecutive lines, then all λ values in consecutive lines.

RESPONSE 18: We made the suggested change.

COMMENT 19: Font size are way too small for proper reading in figures 3, 4, 5, 6.

RESPONSE 19: Thank you! We increased the font size.

COMMENT 20: Figure 3. There is something wrong with the outliers which appear twice in the figure, one black color with jittering, and one blue circle. The jittering is a good practice which increases readability. But in that case, the outlier should also be jittered. Avoid useless color (blue cirle) legend. Instead, use open (non-outlier) and solid (outlier) circles.

RESPONSE 20: Thanks for your feedback. We ended up removing the boxplots and their outliers from the figures and replacing them with symbols for the mean and standard deviation for each population to make the

figures clearer.

COMMENT 21: Lines 357. “however” instead “howeve”

RESPONSE 21: Thank you for catching this! We fixed the typo.

COMMENT 22: Line 507. Italic for Latin name.

RESPONSE 22: Good catch. We fixed it.

Reviewed by Pizza Ka Yee Chow, 26 May 2023 06:42

COMMENT 23: Thank you for inviting me for reviewing this manuscript. The central investigation was examining whether higher behavioural flexibility is seen in the expansion group of grackles that are residing in the edge than those grackles that are locating in the middle. Behavioural flexibility was measured in three ways using reversal learning, innovation, and exploration. The authors found that only innovation and exploration have a difference between the old and new population. Some of the behavioural traits are repeatability.

I find the work interesting and from my personal experience, the work is valuable because it is not easy to conduct such kind of field work; having individuals to complete two or more tasks takes a lot of time, efforts and dedication! The experimental protocol measuring behavioural flexibility is suitable, though many more information are needed to be clarify. The analysis using stimulation models based on previous data or existing literature of another avian species deems appropriate. Excellent on reporting what was proposed in the preregistrations and the current actions.

However, I also find the current manuscript is quite confusing, from concept to terminology and how information is sharing with readers. It definitely needs some rewriting to clarify these confusions, especially shortening sentences will help to make the meaning clearer.

RESPONSE 23: Thank you so much for agreeing to review the Stage 2 after having also reviewed the Stage 1! This continuity is really useful to the pre- and post-study peer review process. And your feedback is super helpful :-) We address your comments on how we can improve the manuscript below. Thank you also for acknowledging how much effort this all took - it certainly was a lot to take on and we are really happy to be at the point where we can learn from our results!

COMMENT 24: Main queries: 1) The central idea is basically testing Wright et al., (2010) adaptive flexibility hypothesis. Why do the authors not cite this hypothesis directly? If I am wrong, I think the authors need to specify how they have revised the hypothesis for the current study.

RESPONSE 24: Funnily enough, we didn’t see Wright et al. (2010) until after we had written the preregistration. We added the Wright citation to the intro at the post-study stage, but it would be dishonest to say that we were explicitly testing the adaptive flexibility hypothesis with this article. We didn’t revise the hypothesis in the post-study article because that is not permitted after the preregistration received in principle acceptance.

COMMENT 25: 2) Measurements: exploration. What latency to approach a novel environment and not how much an individual move within a novel environment an indicator of exploration? Another measurement persistence: there are different ways to measure it, if the author are only giving 5 minutes to a grackle, do the authors really think that it is measuring persistence and not motivation?

RESPONSE 25: Several years ago, Logan (2016) measured exploration as activity level in a grackle’s first hour in the aviaries. However, this seemed to be more a measure of stress then actual exploratory behavior where the

individual is seeking to explore something new without negative (being in captivity for the first time) or positive (having preferred foods next to the object) associations around it. Logan later decided to use Mettke-Hofmann's designs because they disentangle real exploration from activity or stress or neophobia. Additionally, grackles only rarely entered the novel environment, so we would not have a large enough sample size to compare activity levels within the novel environment. See also our Response 38 for more details on exploration and also how we clarified this in the article. Please see our Response 37 for our explanation about the difference between persistence and motivation.

Logan CJ. 2016. Behavioral flexibility in an invasive bird is independent of other behaviors. PeerJ 4:e2215 https://doi.org/10.7717/peerj.2215

Mettke-Hofmann, C., Lorentzen, S., Schlicht, E., Schneider, J., & Werner, F. (2009). Spatial neophilia and spatial neophobia in resident and migratory warblers (sylvia). Ethology, 115(5), 482–492.

COMMENT 26: 3) What is behavioural flexibility afterall? What is the difference between 'behavioural flexibility', 'behaviour' (line 68) and 'behaviour change' (line 60), 'traits' (line 96), 'behavioural traits' (line 438)? Behaviour seems to be a boarder term here that covers behavioural flexibility, innovativeness and exploration. But then in line 43 and 44, behavioural flexibility is also including behavioural flexibility, innovativeness and exploration.....what is the difference?? standardising the term or explain different terms clearer.

RESPONSE 26: We use the terms "behavioral traits" and "behavior" as referring to any behavior. We can see how this is unclear so we now clarify in the Abstract and in the first paragraph of the Introduction (the lines 43-44 that you referred to) what we mean by flexibility, exploration, innovation, and persistence on their first mention, so hopefully this will show how we consider them distinct from each other. For line 68, we clarified that we "aim to compare four behaviors in wild-caught...". For line 438, we removed the word "traits". We additionally clarified the Introduction by adding "To determine whether a behavior (e.g., flexibility, innovativeness, exploration, persistence) is involved in a rapid geographic range expansion", and we added whole sections to explain more about each of the four behaviors we measure. For line 60, we changed this to say "opportunity to assess the role of behavior across their expansion", which should remove the confusion around whether we were referring to flexibility here (we were referring to all of the behaviors we measure). For line 96, we can't change the wording because it is part of the preregistration that received in principle acceptance, but we can add a clarification. We clarified by adding behaviors in parentheses "at least some traits (behaviors) thought to be involved in range expansions".

COMMENT 27: Introduction: Lie 39-43 This sentence is really long, try spilt it into two or more shorter sentences. Behavioural flexibility is a core variable in the manuscript, so it should be introduced clearly to readers.

RESPONSE 27: Thank you. We split it into two sentences as follows: "It is generally thought that behavioral flexibility plays an important role in the ability of a species to rapidly expand their geographic range (Chow et al., 2016; Griffin & Guez, 2014; e.g., Lefebvre et al., 1997; Sol et al., 2002, 2005, 2007; Sol & Lefebvre, 2000). Behavioral flexibility is the ability to change behavior when circumstances change through packaging information and making it available to other cognitive processes (see Mikhalevich et al., 2017 for theoretical background on our flexibility definition)."

COMMENT 28: Line 68, wait, behaviour is a much broader term than behavioural flexibility but interchangeable? And how do these terms different from behaviour change (line 60)

RESPONSE 28: Please see Response 26.

COMMENT 29: Line 95: 'prediction 1' is not necessary because there is only one prediction

RESPONSE 29: We can see where this would be confusing because it is a consequence of splitting the preregistration into multiple post-study articles. We deleted the "1" to reduce this confusion.

COMMENT 30: Lie 96 – trait? There are too many terms that may or may not indicating the same thing now… need clarifications here.

RESPONSE 30: Please see Response 26.

COMMENT 31: Methods: 1) Need more information about why some of the individuals were found at both sites? How close are Woodland and Bufferlands, can these locations be placed on figure 1?

RESPONSE 31: Thanks for asking! We were really excited to learn that they travel this far once we started banding them so it is nice to be able to include more detail in the article. Woodland and the Bufferlands are 32 km apart and we added this to the sentence. We also added an inset to Figure 1 showing a close up of Woodland and the two trap sites.

COMMENT 32: 2) The Authors make a good point of testing certain age group but the choice of age group is not entirely clearly to me as in how testing adults is related to the prediction. Say in my main study species, squirrels, the juveniles disperse the most, and they are (logically) those that show the highest flexibility and explorations, and hence, should test the younger and not the older squirrels – if they are the ones who settle in the edge of the expanding population, right? What about grackles? what is the dispersal characteristics in grackles? Providing more information about the dispersal and how it relates to the hypothesis would be helpful here.

RESPONSE 32: Good point. We originally set out to only test adults because cognition can vary across the developmental period with adults being the most cognitively developed. Therefore, we decided to test adults so we could measure the maximum of what this species is capable of. However, the Arizona grackles were so difficult to catch that we ended up bringing a juvenile into the aviaries to see if his performance was the same as the adults. If so, we could include juveniles as test subjects, which would allow us to more feasibly reach our minimum sample size. The juvenile's performance was well within the range of the adults so we decided that their cognitive abilities must develop pretty early and it should be safe for us to test juveniles. Thereafter, we prioritized trapping adults, but we also brought juveniles into the aviaries if we were struggling to meet our minimum sample size.

To more directly answer your question, movement into new areas appears to be separate from dispersal. Observations from members of the Yolo Audubon Society in Woodland, Davis, and Sacramento suggest that movement into new areas is most likely by adults or groups of mixed age individuals (Yolo Audubon Society's newsletter The Burrowing Owl). This is because the time of year that the first great-tailed grackles were observed on the ground in this county was in late February 2002. This would make the individuals either older juveniles about to enter their first breeding season, or adults. When grackles are in groups, they are usually a mixture of ages and not juvenile flocks. Therefore, it is highly likely that either adults or adults and juveniles move into new areas together. We have also observed at our Woodland site that adults can move up to 32 km twice in a year. This suggests that adults can move long-distances and that this is a separate process from dispersal as a juvenile. We cannot identify dispersers within a given population because the time scale of our DNA data doesn't allow for that high of a resolution, so we won't be able to determine whether the more flexible/exploratory/etc. individuals dispersed farther. We added to Methods > Sample: "Observations from members of the Yolo Audubon Society in Woodland, Davis, and Sacramento, California suggest that movement into new areas is most likely by adults or groups of mixed age individuals (Yolo Audubon Society's newsletter

The Burrowing Owl). Accordingly, if there are differences associated with being at the edge, these differences should also be expressed in adults. ”

COMMENT 33: 3) Figure 2 is shown too early in the manuscript, would be better if it is placed around line 186 when talking about the experimental order. However, the figure is a bit misleading as it seems to indicate all birds went through the same order of the test, but it said the order was counterbalanced for birds. Need some clarifications in the figure captions or use A, B, C to separate the three figures in figure 2.

RESPONSE 33: Thank you! We moved Figure 2 to the bottom of the section and added to the caption: “The order of the flexibility and innovativeness experiments was counterbalanced for the California grackles and they received their first exploration as close as possible to day 8 in the aviaries. The Arizona grackles received the flexibility experiment first (because they underwent a flexibility manipulation) and the innovativeness experiment and exploration assay afterward (note that there could have been other experiments between the flexibility experiment and the innovation experiment and exploration assay because their test battery was much larger than that of the California birds). See the test history for each bird in the gxpopbehaviorhabitatq1_data_testhistory.csv data sheet at @logan2023xpopdata.”

We also added a sentence about this to Methods > Experimental Order, which did not contain the experimental order for the Arizona grackles. The reason was because the Arizona grackles were already undergoing the flexibility manipulation experiment as part of a separate preregistration. When we wrote the current preregistration, we were focusing on the field sites that we would compare with the Arizona grackles because we had not yet collected this data. However, we realize that we should have included information about how the Arizona grackles were treated differently due to their different test battery. We have now corrected this by adding the information here as you suggested.

COMMENT 34: 4) Line 124 ‘adults are identified… adulthood.’ There is a jump between ideas, would be better if this sentence is before ‘however, due to difficulties in trapping this species…’.

RESPONSE 34: Good idea, thanks! We made the change.

COMMENT 35: 5) Line 127-134 ‘mist nets decrease the likelihood of… the fifth day instead)’… this information really should be placed close to the beginning after the sites are mentioned. By the procedure order – you got a site identified – set a new – caught the birds – it’s adult! – put ring on their leg – bring in the lab to habituate the set up.

RESPONSE 35: Great suggestion, thank you so much! We made the change.

COMMENT 36: 6) Innovation: would the authors think that solving more problem = higher exploration too?

RESPONSE 36: No, we define exploration as the willingness to approach novel objects to gain information in the absence of nearby food. Innovation is about whether they are able to discover the methods for accessing food from the 4 differently operated doors on the multiaccess box loci. We added short descriptions of these variables to the abstract to distinguish them from each other.

COMMENT 37: 7) Persistence: persistence comes out of no where, some intro why measuring this trait is important needed. The way the authors measured persistence seemed to be motivation to me… though the two terms are interchangeable.

RESPONSE 37: Good point, sorry about that. We agree that it is difficult to disentangle persistence and

motivation. Indeed, there doesn't seem to be a straightforward way to do this based on our reading of the literature. We try to increase and standardize motivation by removing the maintenance diet and using preferred foods in the tests. Logan previously measured persistence in grackles in Santa Barbara, California as the attempt rate: the number of times they came to the table (on which the test apparatus was placed) or interacted with the apparatus or a stick tool across each bird's 105 min in the experiment (Logan 2016). We don't feel like this persistence measure captured individual variation that we noticed anecdotally. We wanted this measure of persistence to quantify the individual differences that we observe as we interacted with them on a daily basis. We noticed that individuals were definitely unique, but we didn't feel like the existing individual differences measures were able to capture the variation we noticed.

We decided to try the persistence measure here, which is the number of trials participated in divided by the total number of trials offered. We had the intuition that this could potentially be a better measure of persistence because one of the biggest differences we see among the grackles is their willingness to participate in testing sessions, how many trials they want to participate in per day, and how often they have slow days. With this measure, if an individual does not interact with the test, it still accumulates data, unlike many other measures, for example, counting the number of touches to an apparatus or the amount of work time. Our new measure seemed like a better proxy of persistence because it captures all of these elements, which are elements that we perceive as relating to persistence as we work with the individuals. For example, some grackles only complete a few trials per day - they are slow workers, but they continue to participate and complete experiments and this is one type of personality. One such grackle, Ak'xi, has a persistence score of 0.80. Other grackles will participate in as many trials per day as they can get, and this is a different type of personality (they have high persistence scores, around 1.00). One grackle, Tzanatl preciosa, completed all of her experiments and was a pretty active participator through her initial color discrimination, then made the wrong choice on the first trial of her reversal and stopped participating for 2 days, then made another wrong choice and stopped participating for 7 days, then became an active participator again and ended up being the fastest grackle to reverse a preference that we have ever tested. Her persistence score is 0.59. Another grackle, Wachil, almost never participated and also didn't finish many experiments - her persistence score is 0.33. This persistence measure appears to allow us to detect these distinct personalities.

Additionally, there would be no persistence data for the reversal learning test if we used the number of touches to the apparatus because the grackles don't touch the apparatuses, they just touch the food after making a choice. Our measure of persistence allows us to obtain persistence data for experiments that wouldn't otherwise be able to be included. To provide background and clarify our choice of persistence measure, we added the following to the Introduction: "Persistence, "a measure of task-directed motivation" (Griffin & Guez 2014), could facilitate a geographic range expansion because it can improve problem solving success [@morand2011innovators] and efficiency [@chow2016practice]. However, persistence is measured in a variety of ways (e.g., work time, number of touches to the test apparatus, number/frequency of unsuccessful manipulations, etc., see @griffin2014innovation for a review), which makes it difficult to compare across studies. Many measures of persistence are resource intensive to collect because they involve hundreds of hours of video coding, which could prohibit some researchers from being able to measure this variable due to time and financial constraints. Therefore, we developed an easy to calculate measure that we believe better represents task-directed motivation in grackles: the number of trials participated in divided by the total number of trials offered."

Logan CJ. 2016. Behavioral flexibility in an invasive bird is independent of other behaviors. PeerJ 4:e2215 https://doi.org/10.7717/peerj.2215

COMMENT 38: 8) Exploration. The way the authors measured exploration seems to be neophobia

RESPONSE 38: We apologize for not being more descriptive about this test in the methods section. Neophobia is an ambiguous term that has been similarly used to refer to the personality traits of boldness and exploration.

As such, definitions and methods for assessing exploration and boldness are often confused. We conducted a separate experiment (McCune et al. 2019) to explicitly separate these personality traits and validate which methods are appropriate for measuring each trait in our system (Carter et al. 2013). Briefly: we followed the methodology of Mettke-Hofmann et al. (2009) for measuring exploration, which distinguishes exploration from boldness (which they call neophobia) in the following way: neophobia is the willingness to eat food next to a novel object, whereas exploration is the willingness to approach a novel object in the absence of food. The latter is an indication of how willing they are to explore the novel object without needing to for biological necessities like food, water or shelter. Furthermore, Reale et al. (2007) define boldness as the behavioral response to a threat, whereas exploration is the propensity to investigate novelty.

Based on these definitions, we used multiple methods to validate which methods resulted in repeatable performance across time, and that variation in performance was in line with our definitions of exploration and boldness (McCune et al. 2019). Of these multiple methods, we found that performance was repeatable in the exploration of a novel environment (a small tent) within the familiar environment. Furthermore, we found no significant difference in the behavior of grackles in the control condition, where the tent was absent, versus the experimental condition with the tent, indicating they did not perceive the tent as threatening. This was in contrast to the other methods that included potentially threatening objects where grackle performance was not repeatable across time and grackles avoided the objects in the experimental conditions compared to the control conditions.

We added a sentence to the methods section to briefly refer to our validation of this method as a measure of an exploratory trait rather than neophobia: "In a previous experiment [@mccune2019exploration], we validated that grackles did not perceive the tent as threatening (i.e., it was not a measure of boldness)."

Carter, A., Feeney, W., Marshall, H., Cowlishaw, G., & Heinsohn, R. (2013). Animal personality: What are behavioural ecologists measuring? Biological Reviews, 88(2), 465-475.

McCune, KB, MacPherson, M, Rowney, C, Bergeron, L, Folsom, M, & Logan, C. (2019). Is behavioral flexibility linked with exploration, but not boldness, persistence, or motor diversity? In Principle Acceptance by PCI Ecology of the Version on 27 Mar 2019. http://corinalogan.com/Preregistrations/g_exploration.html

Mettke-Hofmann, C., Lorentzen, S., Schlicht, E., Schneider, J., & Werner, F. (2009). Spatial neophilia and spatial neophobia in resident and migratory warblers (sylvia). Ethology, 115(5), 482–492.

Reale, D., Reader, S., Sol, D., McDougall, P., & Dingemanse, N. (2007). Integrating animal temperament within ecology and evolution. Biological Reviews, 82(2), 291-318.

COMMENT 39: 9) Line 179: what is this motivation test entails?

RESPONSE 39: Great question. We clarified as follows (Methods > Protocols > Exploration): "and we first conducted a motivation test where we placed a piece of preferred food on the ground and waited out of view for 5 min. We only proceeded with the exploration assay if the bird ate the food. This motivation test allowed us to determine whether the grackle is interested in coming to the ground at all, where for example, a grackle might not eat the food because it has just bathed and is primarily focused on preening and drying feathers."

COMMENT 40: Discussion. Line 432-435 'our measures of flexibility.... Meaningfully compared' – repeatability is an additional and informative analysis beside testing the major hypothesis. These should be placed after answering the major research question.

RESPONSE 40: Thank you, we moved this sentence to the end of the paragraph.

COMMENT 41: Line 438. Another term here that may be indicating the same or different thing of 'behavioural flexibility'...

RESPONSE 41: “behavioral traits” refers to any behavior. Please see Responses 26 and 28 for how we clarified this in the Introduction.

COMMENT 42: Line 440. The authors may want to reclarify what this ‘flexibility’ is and how it was measured…..

RESPONSE 42: Good point. We added to the sentence (Discussion): “We found no support for the hypothesis that a higher average flexibility (reversal learning of a color preference) is required in an edge population. That flexibility, the ability to change behavior in reaction to changing circumstances through packaging information and making it available to other cognitive processes, was not on average higher among individuals at the edge of the expansion range indicates that flexibility is not a latent trait that is called upon when individuals move into new areas”

COMMENT 43: Line 487. I agree with this possibility… the test may be conducted ‘too late’ (of grackels point of view and species characteristics). May worth adding more information the population that is in this study and suggest what would be a better idea to test this hypothesis again (what is ‘long enough’ line 489?).

RESPONSE 43: The problem is that it wasn’t feasible for us to keep the aviaries in one location and conduct the tests that we did in a more northerly population because those populations are too small or too ephemeral. To be able to move the aviaries from place to place to catch a few grackles across multiple locations closer to the edge would have required a much larger time and financial investment. We added a sentence in this paragraph to clarify that we were as far on the edge as we could get in order to reach our minimum sample size: “ Because the more northern populations are small and ephemeral, to obtain our minimum sample sizes, a different and more geographically expansive experimental approach would be necessary. Future efforts could focus on a broader geographic area across Washington or Oregon for capturing these individuals to measure flexibility and other behaviors to add important information to our understanding of the relationship between variation in behavior and the ability of species to expand their range”.

Reviewed by anonymous reviewer, 24 May 2023 10:06
COMMENT 44: This is a very interesting study that examines the role of behavior, specifically behavioral flexibility and innovation, in the expansion of a species. The study aims to investigate whether individuals at the periphery of a population, in newly colonized areas, exhibit novel behaviors or rather intensify certain behaviors already present in the entire population. Overall, the findings suggest that individuals in newer populations demonstrate greater innovation but engage in less exploration. However, no differences in flexibility or persistence were observed. Despite the interesting results, I have several general and specific considerations that could enhance the organization and comprehensibility of the study.

RESPONSE 44: Thank you very much for sharing your comments on how we can improve the manuscript!

COMMENT 45: General considerations. On the one hand, the introduction seems too specific, I would like to see what has been studied and what behavioral variations have been found in different species. Also, I would not include the part of the great-tailed grackles in the introduction, but I would make a specific section in material and methods on the study species. In addition, the objective of the article is mentioned from the beginning, when this should only be specified at the end of the section (i.e., introduction). In the previous paragraphs, it should be mentioned which are the lack of knowledge in the subject matter.

RESPONSE 45: Thank you for this feedback. The introduction in the preregistration was very general because multiple articles are coming from the one preregistration. So this is a great opportunity for us to make this introduction more specific to this article. We expanded the introduction to discuss how each of the four

behaviors is expected to relate to geographic range expansions and we now go into more detail about each of the behaviors, and point out the holes in the literature. We think it is important to keep the great-tailed grackle section in the introduction because it provides the background for the study and links it with the theory and examples that we now added.

COMMENT 46: On the other hand, I consider the material and methods section to be very complex and a bit cluttered. I consider that material and methods need to be restructured, simplifying the sections that this section is divided into and explaining in greater depth and with greater order all the experiments and the decisions taken with the individuals handled. I understand that analysis and evaluation are not easy, but I think it is important to follow the workflow.

RESPONSE 46: Please see Response 4 for how we clarified the methods by better relating them with the supplementary material, as well as an explanation for why so much of the methods are in the supplementary material. See Response 72 for how we simplified the sections, and Responses 33 and 48 for how we expand upon the order of the experiments.

COMMENT 47: One of my main doubts about the experiments is to know if you have considered that being a social species, learning by imitation/observation (i.e., use of social information) has an important role, which however in your experiments you are not considering it.

RESPONSE 47: In the current article, it is important to measure each individual's baseline cognitive ability, which needs to be tested in visual isolation of other individuals so we can ensure that they aren't learning about the test from each other. This is standard practice for comparative cognition experiments. For example, obtaining an individual's exploratory behavior in the wild can be confounded because a dominant might displace them and exclude others from the novel environment, which would make the subordinate individual's exploration latency much longer than it normally would be for this individual if it was tested alone. For research questions that involve a social component, we have a separate preregistration that has received in principle acceptance at PCI Ecology on whether this species socially learns and, if so, through what mechanisms (McCune et al. 2019). We are still in the process of collecting the social learning data so we don't yet have results.

McCune KB, McElreath R, Logan CJ. 2019. Investigating the use of learning mechanisms in a species that is rapidly expanding its geographic range. (http://corinalogan.com/Preregistrations/g_sociallearning.html). In principle acceptance by PCI Ecology of the version on 11 Oct 2019

COMMENT 48: Also, in the behavioral flexibility analyses you include the learning rate, however I do not see that you include the option of learning or behavioral change in the rest of the experiments. Not only learning but adaptation to captivity and to the experiments themselves should be considered, and this would be easily modeled by including the number of experiments the individual has previously undergone, or the number of days in captivity (which also serves as a proxy for experience with the experiments).

RESPONSE 48: The number of days in captivity might not be a relevant proxy for how many experiments an individual has experienced because some birds participate in only a few trials per day and do not complete experiments very quickly. We ensured that their habituation to captivity and to the experimental apparatuses did not interfere with testing their cognitive abilities on either of the two experiments (flexibility and innovation) by making sure each bird passed their habituation criterion for each experiment before they were able to move on to test trials (see Supplementary Material 5 for more details).

The learning rate is one component of the two components specifically quantified from the performance on the flexibility (color tube reversal) experiment. It is therefore only relevant to this flexibility construct and would not be appropriate to apply it to other contexts. We do consider the effect of flexible behavior on interactions

with a multiaccess box in another manuscript (Logan et al. 2023). But in this manuscript we only consider the one test of flexibility in which the circumstances changed and the birds had to change their behavior. The innovativeness experiment with the multiaccess box, only measured how many loci they solved and not how they changed their behavior in response to changes in the box.

There is an issue that we failed to recognize until your comment and that is the influence of a flexibility manipulation we conducted on 8 of the Arizona grackles. Since submitting this article for review, we obtained results from a different article on the Arizona grackles (Logan et al. 2023). The other article was a flexibility manipulation, which involved conducting serial reversals with the birds in the manipulated condition until they consistently passed each reversal quickly, while the control birds received only one reversal and a similar number of trials with identically colored control tubes where both contained food. We found that the manipulated birds performed better on the innovation experiment than control birds, which indicates that specific experience is necessary for birds to improve their performance, rather than time in captivity or experience with experiments. The reversal learning test in the Arizona population was always conducted before the innovativeness and exploration tests because we were determining whether the manipulation affected performance on these other tests. Because the manipulation causally influenced their innovativeness (we are not yet sure about their exploration or persistence because we are still conducting these analyses), we now need to assume that the manipulation could be biasing the results in the current article (e.g., we made some of the Arizona grackles faster, which could bring down the Arizona average and thus find no difference between the populations if the California grackles are potentially faster after accounting for the manipulated individuals). Therefore, we added Treatment (whether an AZ grackle was manipulated or not) as a variable in the models to account for this. We reran the models and the power analyses and found that two of the four results changed: there is now no difference in innovativeness or exploration between populations, and there is a difference in persistence with the edge population being more persistent. The flexibility result stayed the same: there is still no difference in average flexibility levels between populations.

Because we added the treatment term to the models, we reran the power analyses to determine whether adding this term caused any changes in the minimum sample sizes. The minimum sample sizes either did not change, they got smaller: the minimum sample size for persistence went from 18 to 15 (because we got better at making our simulations more accurate), or they stayed the same and the number of seconds by which the two sites must differ increased (exploration). We could not rerun the flexibility power analysis with treatment added to the model because the Santa Barbara grackle data were used for the power analysis and these birds did not have a flexibility manipulation. Instead, we reran the flexibility analysis in the Results section using data from only the first reversals. This made it so treatment is not a variable that needs to be included in the model and it made our results relevant to the flexibility power analysis. The result stayed the same: there were no site differences in flexibility (phi or lambda).

Logan CJ, Lukas D, Blaisdell AP, Johnson-Ulrich Z, MacPherson M, Seitz BM, Sevchik A, McCuneKB. 2023. Behavioral flexibility is manipulable and it improves flexibility and innovativeness in a new context. EcoEvoRxiv, version 5, peer reviewed and recommended by Peer Community in Ecology. doi: https://doi.org/10.32942/osf.io/5z8xs

COMMENT 49: Specific comments. Introduction. I believe that a more general approach to the central issue is needed: to establish the importance of behavioral flexibility in species expansion. Has it been studied in other taxa: mammals, for example? And in other species?

RESPONSE 49: Thank you for this feedback. We have revised the Introduction accordingly (also see Response 45).

COMMENT 50: The more specific part of the introduction where the species is discussed as a specie of interest (lines 56-67), I would move it to a specific section of material and methods called ”Study species”.

RESPONSE 50: Please see Response 45.

COMMENT 51: Lines 43-44. You talk about innovation, but you do not define the concept. Also learning methods (individual or social) influence behavioral flexibility, perhaps it is important to mention it.

RESPONSE 51: We now expand on innovation, its definition, and how it relates to flexibility and a range expansion in the Introduction. We addressed why we chose an asocial context in the Introduction as follows: "While it is possible for individuals in the wild to learn asocially and socially about new foods or foraging techniques to assess whether the risks are low enough to encourage exploration behavior, we focused on measuring these four behaviors in an asocial context to allow us to obtain the individual's actual cognitive performance (i.e., in the absence of dominant individuals who might hinder subordinates from participating)."

COMMENT 52: Lines 48-49. In these sentences you talk about what is expected, I would leave this for the end of the introduction, where the hypotheses and predictions are established. This paragraph lacks a more general perspective on the role of behavior in population dynamics.

RESPONSE 52: We have now greatly expanded the Introduction and we moved this sentence to later in the Introduction where we start to expand on our study.

COMMENT 53: Lines 64-67. Have neophobia levels been studied in this species, and are they lower than those of other species with other ecological requirements?

RESPONSE 53: We investigated neophobia in this species in Santa Barbara (Logan 2016) and in Arizona (McCune et al. 2019), however neophobia was not a behavior we tested in the current article (see Response 38). Therefore, we prefer not to get into a discussion around neophobia in this article so as to keep our message clearer. Also, it is very difficult to compare levels of neophobia across studies because it is often measured and/or calculated differently.

Logan CJ. 2016. Behavioral flexibility in an invasive bird is independent of other behaviors. PeerJ 4:e2215 https://doi.org/10.7717/peerj.2215

McCune, KB, MacPherson, M, Rowney, C, Bergeron, L, Folsom, M, & Logan, C. (2019). Is behavioral flexibility linked with exploration, but not boldness, persistence, or motor diversity? In Principle Acceptance by PCI Ecology of the Version on 27 Mar 2019. http://corinalogan.com/Preregistrations/g_exploration.html

COMMENT 54: Lines 71-72. Why are you looking at variances? Specify.

RESPONSE 54: If edge habitats have different foraging niches than the habitats from which the individuals moved, there are two ways these new foraging niches could be exploited. The first is that all individuals have high flexibility/exploration/innovativeness/persistence, which means that all individuals can exploit new foraging niches. The alternative is that these new foraging niches are discovered by some individuals who have high flexibility/exploration/innovativeness/persistence, from whom other individuals might be able to socially learn over time. In this second case, we would expect the variance among individuals to be higher in the edge populations. This is clarified in the Research Question section and we now bolded the words "averages" and "variances" to make them easier to find.

COMMENT 55: Lines 75-76. Is scanning then a measure of time?

RESPONSE 55: Sorry, but we don't understand what is meant by "scanning". Perhaps you mean latency? If so,

latency is the number of seconds it took a bird to approach to within 20 cm of a novel environment. We added "(in seconds)" to the sentence to clarify.

COMMENT 56: Table 1. How is the generation time calculated?

RESPONSE 56: The number of generations at a site is based on a generation length of 5.6 years for this species, which we obtained from BirdLife International (2018). We now include details about how we make the calculation in the Table 1 caption: "The average number of generations was calculated using the number of years of breeding (the "Breeding since" year up to 2020, the final year of data collection in Tempe, and 2022, the final year of data collection in Woodland) divided by the 5.6 year generation length."

BirdLife_International. (2018). Quiscalus mexicanus. The IUCN Red List of Threatened Species 2018, e.T22724308A132174807. http://dx.doi.org/10.2305/IUCN.UK.2018-2.RLTS.T22724308A132174807.en

COMMENT 57: Could it be that the scan is different as a function of time or the number of experiments that have been performed on the individual? In the end, all parameters can be affected by the experience gained.

RESPONSE 57: Please see our Response 48 for details.

COMMENT 58: Lines 92-111. As I have already mentioned in the general comments, I think it is necessary to combine in a single paragraph the main objective of the work, then describe very concisely how it is carried out and the specific hypotheses and predictions.

RESPONSE 58: Please see Response 45 for how we made the Introduction more general and how the four behaviors are measured. Here, the research question and prediction must be kept separate and unchanged to preserve the preregistered format that received in principle acceptance.

COMMENT 59: Line 103-106. Try to be more direct with sentences, rather than starting them with "Perhaps in newly....." Say: "In newly established populations, individuals may need to learn about innovate new foraging techniques or find new food sources."

RESPONSE 59: This is a good suggestion, but we are not able to change specific wording in the Hypothesis section because it was preapproved as part of the in principle acceptance.

COMMENT 60: Line 104. The "and" is not necessary.

RESPONSE 60: We are not seeing which "and" you are referring to in this sentence: "Perhaps in newly established populations, individuals need to learn about and innovate new foraging techniques or find new food sources". The "and" here is necessary because there is a difference between learning about foraging techniques and innovating a new foraging technique. Though, this text was part of the preregistration that received in principle recommendation and we need to keep it as it is because it was preapproved.

COMMENT 61: Line 107. "Higher variances in behavioral traits will indicate that there is..."

RESPONSE 61: Changed as requested.

COMMENT 62: Methods. The methods must be written in the past tense.

RESPONSE 62: We were trying out writing the whole article in present tense as much as possible because our

target journal is Peer Community Journal, which does not have tense requirements. However, either we failed at implementing this effectively or it is something that is too new to be widely accepted, therefore we changed the tense in the methods to past tense.

COMMENT 63: What is the basis for the selection of 3 to 4 days of habituation?

RESPONSE 63: This was the number of days Logan previously used with grackles in Santa Barbara because it gave them some time to adjust to the aviaries and the presence of an experimenter if they needed that amount of time. Birds differ in how long it takes them to adjust, but if a bird is pretty bold, 3 days is more than enough. For those birds who need longer to adjust, they get as much habituation time as they need until they are ready to start participating. If they don't habituate in a reasonable amount of time (it depends on the bird's behavior), we release them back to the wild.

COMMENT 64: The fact that it is a sociable species and that you put them individually in the aviary may not be modifying their behavior?

RESPONSE 64: We are interested in an individual's inherent cognitive abilities and behavior for these tests, which is why we test them in visual isolation of each other. It eliminates the confounds about what they might be learning about the test from social learning before and as they interact with the apparatus themselves. Testing them in visual isolation of conspecifics allows for testing a wider variation in individual responses (i.e., across all dominance ranks) because if individuals were not isolated then dominance interactions, and not individual capabilities, may determine who participates and this could potentially minimize the variation we are able to measure within the population and bias our sample. While great-tailed grackles are social, they aren't as affiliative as monogamous bird species are and they don't get very close to each other in the wild (e.g., our years of focal follows showed very few close proximity behaviors when birds were within one body length of each other). It is possible that individual captivity modified the grackle's behavior, but our main objective was to compare behavior across populations, and because both populations received the same methods, it is unlikely that this affected our results. Several years ago, Logan attempted an experiment with grackles in Santa Barbara that required 4 grackles to be in the same aviary, but they were entirely focused on dominance interactions, which appeared stressful so they were separated and kept separate. Therefore, it is for their safety that they are housed individually.

COMMENT 65: It is not clear how many individuals from each population are used, I recommend specifying this clearly in the text.

RESPONSE 65: Great point! We updated the Results sections to include the specific sample size for each population for each test. Sample sizes varied according to test, which is why we prefer to report the population sample sizes here rather than in the Methods.

COMMENT 66: Were there individuals that spent different lengths of time in captivity? That could influence the outcome of the experiments.

RESPONSE 66: Yes, individuals spent different amounts of time in the aviaries - it depended on how fast they got through their test battery. When they finished, they were released back to the wild. Please see our Response 48 for more information.

COMMENT 67: Lines 135-137. I consider that it is not necessary to explain why field work was not carried out in other areas, although the arguments are weighty, what is important is the data obtained and analyzed.

RESPONSE 67: We moved this piece to the section “Post-study choices made since receiving in principle recommendation” because it is important to document all of the differences.

COMMENT 68: Line 151. Regarding the flexibility protocol, do all individuals require the same number of attempts to fix the selection of a color?

RESPONSE 68: There is a lot of individual variation in how many trials it takes a bird to pass our criterion of 17 out of 20 trials correct on the initial color discrimination and then to reverse this color preference. We now present each bird’s summary results in the new Table 2 at the top of the Results section.

COMMENT 69: Line 146. The two lines corresponding to the ”Data collection stopping rule” section would be included in the previous section.

RESPONSE 69: We made the change.

COMMENT 70: Line 171. In line with another of my previous comments: when do you consider that an individual is habituated? Are there differences between individuals in the habituation process?

RESPONSE 70: There are individual differences in how long it takes them to habituate to the aviaries and the testing process. The first thing a bird experiences in the aviaries is a habituation test that differs depending on what their first experiment is. We see whether they are willing to interact with the habituation object in an experimental setting (i.e., when an experimenter is sitting in the aviary aisle). We consider a bird is habituated when they quickly interact with the object without exhibiting behaviors indicative of fear (approaching slowly or circling the object, moving quickly or jumping away after interacting, etc.). They must pass this habituation criterion to be able to move on to a test. We updated this sentence to point readers to Supplementary Material 5 where the habituation protocols are detailed. The only exception to this rule that birds must first pass the habituation test before moving on to other experiments was in the California grackles where we still gave grackles that did not habituate the exploration assay because it did not involve an experimenter sitting in view or require the birds to interact with any apparatus. Whereas, in Arizona, the birds were not able to move on to any other test until they finished their reversal learning experiment because our main goal in that population was to conduct a flexibility manipulation and determine the influence on subsequent experiments.

COMMENT 71: Lines 176-178. I do not understand the definition of Time 1 and Time 2. Rewrite.

RESPONSE 71: Sorry for the confusion. We revised the sentence to say (Methods > Protocols): “Exploration was measured as the latency to approach within 20 cm of a novel environment inside of their familiar aviary environment and this test was conducted two times for each bird so we can obtain individual consistency measures. Time 1 occurs on the individual’s 8th day in the aviary and Time 2 occurs 1 week after Time 1.”

COMMENT 72: The section on Randomization and counterbalancing, together with that on experimental order, I would include it in the previous sections, simply adding the sentences in this regard (without titles).

RESPONSE 72: We made the change. We also removed the Open Data section and moved it to the end of the manuscript (before the Ethics section) because this follows the final PCI Ecology formatting.

COMMENT 73: Line 196. I would call the general section: statistical analyses. The other sections that follow, e.g., ”flexibility analyses” or ”model and simulation” should be subsections of this one.

RESPONSE 73: We changed the heading title as suggested. The following subheadings (e.g., Flexibility analyses) are marked as subsections.

COMMENT 74: Lines 199-201. Wouldn't this be a good opportunity to test whether the experience of participating in different experiments influences the individual's behavioral skills? I think it may be important to take into account how many experiments you have participated in and how many different types, as this can be a learning experience for the individual.

RESPONSE 74: Please see our Response 48.

COMMENT 75: Line 242. The version of R is already specified above, remove in one of the two places.

RESPONSE 75: Thank you, we removed it here.

COMMENT 76: Line 214. Can you define what the attraction score is?

RESPONSE 76: Good point. We added clarifications as follows (Methods > Statistical analyses > Flexibility analyses): "The attraction score is the weight an individual gives to a particular option based on its past reward history for that option with attractions increasing if they received a reward when previously choosing that option. The decision regarding which of the two options to make is determined by the relative weights of the two attraction scores (each option gets its own attraction score)"

COMMENT 77: Line 215. Define "rate of deviating from learned attractions".

RESPONSE 77: Please see Response 76.

COMMENT 78: Results. Figure 3. Did the value of learning rate change as the individual was subjected to more experiments? Perhaps this learning rate increases to a greater extent in individuals belonging to new populations.

RESPONSE 78: Please see Response 48.

COMMENT 79: Discussion. In general, as in the case of the introduction, I consider that the discussion lacks more general references.

RESPONSE 79: Thank you for pointing this out. We now added general references throughout the discussion.

COMMENT 80: Line 445. This paragraph is totally linked to the previous one. Join.

RESPONSE 80: We made the change.

COMMENT 81: Actually, behavioral flexibility may be playing a fundamental role in both populations, as this species is so closely linked to humanized environments. Perhaps species from more stable, more natural environments (e.g., forested environments) show lower levels of flexibility? Haven't similar species living in different environments been compared (they don't have to be birds)?

RESPONSE 81: Thank you for bringing this up. We agree that flexibility may be playing a role in both populations

because this species is highly flexible compared with many other species. We updated the Discussion to reflect this: "We found no support for the hypothesis that a higher average flexibility is required in an edge population". Flexibility measured in a comparable way through reversal learning has been conducted in many species. However, we were not able to find reversal learning studies of closely related species who live in different environments. When there were studies of closely related species (who were not domesticated or living permanently in captivity), they were living in the same environment. For example, great tits and blue tits are flexible and live in forests (Morand-Ferron et al. 2022), the three species of Darwin's finches are similarly flexible and live in natural habitats (Tebbich et al. 2010), the three corvid species (pinyon jays, Clark's nutcrackers, and California scrub jays) are not as flexible as grackles (Bond et al. 2007) and they live in forests except for the California scrub jay who additionally lives in suburban areas. So current research indicates there is no relationship between flexibility and living in human-modified environments.

Bond, A., Kamil, A., & Balda, R. (2007). Serial Reversal Learning and the Evolution of Behavioral Flexibility in Three Species of North American Corvids (Gymnorhinus cyanocephalus, Nucifraga columbiana, Aphelocoma californica). Journal of Comparative Psychology 121(4), 372-379.

Morand-Ferron, J., Reichert, M. S., & Quinn, J. L. (2022). Cognitive flexibility in the wild: Individual differences in reversal learning are explained primarily by proactive interference, not by sampling strategies, in two passerine bird species. Learning & Behavior, 50(1), 153-166.

Tebbich, S., Sterelny, K., & Teschke, I. (2010). The tale of the finch: Adaptive radiation and behavioural flexibility. Philosophical Transactions of the Royal Society of London B: Biological Sciences, 365(1543), 1099–1109. https://doi.org/10.1098/rstb.2009.0291

COMMENT 82: Line 476. "great-tailed grackles are flexible on the reversal..." capitalize the first word.

RESPONSE 82: Thank you for catching this! We fixed it.

COMMENT 83: Line 487. Is this population still expanding?

RESPONSE 83: Great question. It is unknown whether their overall numbers in these locations are still increasing because we are the first to individually identify them. We added to the Discussion: "Though it seems that this population is still becoming established, in that they are not found at the Woodland trap site year-round and some individuals at the Sacramento trap site also disappear and reappear for parts of the year."

COMMENT 84: Lines 495-497. Why, if there are populations that are still establishing, were individuals from these areas not used?

RESPONSE 84: Great point. We added to the Discussion: "Because the more northern populations are still small and ephemeral, to obtain our minimum sample sizes, a different and more geographically expansive experimental approach would be necessary. Future efforts could focus on a broader geographic area across Washington or Oregon for capturing these individuals to measure flexibility and other behaviors to add important information to our understanding of the relationship between variation in behavior and the ability of species to expand their range."

COMMENT 85: LIne 503. It is not clear to me why longer settled populations are more explorative. Earlier you say that in guinea pigs no changes in this type of behavior have been seen.

RESPONSE 85: As we discovered in Response 48, this result has now changed and now there are no population differences in exploration. So we updated this part of the Discussion accordingly.

COMMENT 86: Lines 524-527. I don't understand the difference between flexibility and innovation. One seems to be directly related to the other.

RESPONSE 86: Please see our Response 26 for a description of how we revised the introduction to better define our terms from the beginning.

Decision by Esther Sebastián González, posted 30 May 2023, validated 30 May 2023

Preprint merits a revision

Dear Dr. Corina Logan,

Thank you very much for your preprint. It has been reviewed by three experienced researchers and they all have underlined the importance of your work. However, they also see many points that need to be improved, especially to clarify terminology and methodology. The three reviewers have made an excellent job and have provided with detailed suggestions and discussion points that I think will largely improve the quality of your study. Because of the many issues raised, I cannot recommend your preprint in its present form, but I will be happy to read a revised version of it that takes into account or discusses all the points raised by the reviewers. I am looking forward to read this revised version.

Regards,
Esther

Reviewed by Francois-Xavier Dechaume-Moncharmont, 29 May 2023

This article presents the results from a preregistered study on cognitive ecology in Grackle, from two populations in the USA, one being in the border of the expansion range of the species. Most notably, the objectives of the study were to assess the differences not only in mean but also in variance in cognitive performances between these two populations, in order to test the hypothesis of a link between invasibility and cognitive traits. The aim and scope of the MS were clear. The experimental design was appropriate and generally well presented. One of the salient features of this study is that it is based on a thorough a priori analysis of statistical power. The authors very rarely deviated from their pre-recorded protocols, and when they did, they stated so unambiguously. My overall impression about this manuscript is positive, and I recommend acceptation after revisions. Indeed, some points deserve clarification prior final acceptance.

Major comments

Writing style. As a non-native English-speaking writer, I won't venture to criticize English or the syntax. Yet, the MS constantly alternates between past and present tense, sometimes in the same paragraph or even in the same sentence ("we investigate… we aim for… we expected… Grackles are habituated… The doors were locked open…"). I strongly recommend rewriting the text in the past tense, so as to present the study and the results as a report on experiments that have already been carried out. Avoid the present tense, which should be confined to expressing general truths. The constant switching between present and past tense gives the very unpleasant impression of reading a MS hastily rewritten from the preregistered protocol.

The statistical analyses (methods and results) are particularly difficult to follow due to several issues. Some important analyses can only be understood when reading supplementary materials, and this information is not cited in the main text. For instance, the authors "estimated 89% compatibility intervals" (lines 358) without explanation. The reader must wait until supplementary materials (lines 604) to find a reference to Richard McElreath's book. The MS should be edited in order to present all the information necessary for the understanding or straightforward references to the corresponding section in supplementary materials.

I also strongly recommend rewriting all the main text, tables and figures in order to avoid highly confounding mathematical notations. λ is used for both the deviation rate in flexibility analysis and a probability of approaching a novel environment in exploration analysis. ϕ is used for both the learning rate in flexibility

analysis and the dispersion rate across bird in exploration analysis. $alpha$ in several model with different distribution law. The author acknowledges that confusion in the middle of the results section (lines 394-395). But that is clearly insufficient. The reader must be guided, and all possible elements of confusion must be eliminated, even is this thorough rewriting process is time consuming. Your time is as precious as the reader's. You write a paper because you are expecting to be read and understood (and possibly cited). A scientific article is not supposed to be a detective story for the reader. The acrimony of this comment reflects the difficulty I had in reading the MS.

I acknowledge that the authors took great care in reporting effect sizes with compatibility intervals. This is unquestionably an excellent practice. Yet, these metrics alone are extremely hard to appreciate. For instance, the authors explain that "The flexibility measure is how many trials it takes them to reverse their color preference" (lines 160). This definition is straightforward and easy to understand from a biological perspective. But the flexibility is later only reported in terms of attraction score, learning rate (ϕ) and rate of deviating from learned attractions (λ). The estimated values of these metrics must certainly be reported in the MS, but the authors should not forget to report and illustrate biologically meaning metrics (so called "how many trials it takes them to reverse preference"). In addition, it is very important to explain in detail what this effect size indices mean, particularly when there are not trivial dependencies between them: "higher ϕ values meant lower λ values" (lines 234).

I took time to careful read the MS and properly think about the methods, yet I still do not understand the analysis reported lines 374, about the innovation. The main issue come from the sentence: "the compatibility interval did not cross zero". As stated lines 246, the response value was to be found in the set \0, 1, 2, 3, 4\. Assuming random distribution of this measure, with most classical distributions, the difference between the average responses of the two populations, should not overlap zero. The null hypothesis should not be zero. But, here, I suspect that I misunderstood something. I recommend that the authors elaborate on the methodology. For instance, how did they compare the two populations? Did they use resampling (bootstrap) methods? If so, provide details about the procedure.

Minor comments

Lines 80 and followings. "Our results DEMONSTRATE that the rapid geographic range expansion of great-tailed grackles is ASSOCIATED WITH individuals differentially expressing particular behaviors in the edge compared to the older population." Without speaking of the epistemological concerns about the fact that Science does not aim at demonstrating anything (such strong statements should be avoided), the study is purely correlative. The FUNCTIONAL link between cognitive skills and dispersion remains to be established. For instance, the difference could come from difference in ecological constraints in the two sites without any connection with the expansion history of the species. The study is sufficiently solid and new in itself, there is no need to overestimate it. I recommend that the authors reformulate the presentation of their results in a more neutral way.

Fig. 2. The far-right drawing (about exploration) is very difficult to understand. Please redraw and give additional explanations in the legend.

Lines 113 and following. In this paragraph, I would greatly appreciate finding information about size, mass and body conditions (Peig & Green 2009) of the birds from the two populations. Difference in body conditions would be particularly important when discussing differences in ecological success in between the two populations.

Lines 150. The bullet points are useless here. They can simply be presented as plain sentences in regular paragraphs.

Lines 150. Avoid hypertext link in a MS. Beside, the protocol is crucial and must be included in the article, possibly as supplementary material, and not as an unreliable googledoc…

Throughout the MS, the reference "Logan et al. 2023" is ambiguous since there are two possible papers corresponding to this abbreviation in the reference section. The main text should be thoroughly edited.

Lines 197 and following. The authors provide a long inventory of the R packages used (psych, irr, rethinking,

rstan, knitr, dplyr, tidyr, cmdstanr, DHARMa, lme4, Rcpp), without the reader knowing for what use and for what precise analysis. Be more specific and give sufficient information about the most important packages. Some accessory packages are clearly less critical (knitr, dplyr, tidyr) because they are mainly used to format data or code. They can be omitted from the main text and only quoted in the source code if they are relevant. In addition, I am surprised by the largely outdated reference (2017) for R language while the cited version of the software (4.1.2) was published in November 2021.

Line 239, Table 1, Fig 3, Table 4, etc. Use consistent notation (either the Greek letter or its transcription in Latin alphabet, e.g. phi, lambda) throughout the MS.

Lines 242. This information about R version is a repetition of information already presented on the previous page.

Lines 316. Remove useless brackets.

Table 2. Since the analyse of the study was to compare the value and their compatibility interval, I recommend to present all ϕ values in consecutive lines, then all λ values in consecutive lines.

Font size are way too small for proper reading in figures 3, 4, 5, 6.

Figure 3. There is something wrong with the outliers which appear twice in the figure, one black color with jittering, and one blue circle. The jittering is a good practice which increases readability. But in that case, the outlier should also be jittered. Avoid useless color (blue cirle) legend. Instead, use open (non-outlier) and solid (outlier) circles.

Lines 357. “however” instead “howeve”

Line 507. Italic for Latin name.

Reviewed by **Pizza Ka Yee Chow**, 26 May 2023

Thank you for inviting me for reviewing this manuscript. The central investigation was examining whether higher behavioural flexibility is seen in the expansion group of grackles that are residing in the edge than those grackles that are locating in the middle. Behavioural flexibility was measured in three ways using reversal learning, innovation, and exploration. The authors found that only innovation and exploration have a difference between the old and new population. Some of the behavioural traits are repeatability.

I find the work interesting and from my personal experience, the work is valuable because it is not easy to conduct such kind of field work; having individuals to complete two or more tasks takes a lot of time, efforts and dedication! The experimental protocol measuring behavioural flexibility is suitable, though many more information are needed to be clarify. The analysis using stimulation models based on previous data or existing literature of another avian species deems appropriate. Excellent on reporting what was proposed in the preregistrations and the current actions.

However, I also find the current manuscript is quite confusing, from concept to terminology and how information is sharing with readers. It definitely needs some rewriting to clarify these confusions, especially shortening sentences will help to make the meaning clearer.

Main queries:

1) The central idea is basically testing Wright et al., (2010) adaptive flexibility hypothesis. Why do the authors not cite this hypothesis directly? If I am wrong, I think the authors need to specify how they have revised the hypothesis for the current study.

2) Measurements: exploration. What latency to approach a novel environment and not how much an individual move within a novel environment an indicator of exploration? Another measurement persistence: there are different ways to measure it, if the author are only giving 5 minutes to a grackle, do the authors really think that it is measuring persistence and not motivation?

3) What is behavioural flexibility afterall? What is the difference between ‘behavioural flexibility’, ‘behaviour’ (line 68) and ‘behaviour change’ (line 60), ‘traits’ (line 96), ‘behavioural traits’ (line 438)? Behaviour seems to be a boarder term here that covers behavioural flexibility, innovativeness and exploration. But then in line 43 and

44, behavioural flexibility is also including behavioural flexibility, innovativeness and exploration.....what is the difference?? standardising the term or explain different terms clearer.

Introduction:

Lie 39-43 This sentence is really long, try spilt it into two or more shorter sentences. Behavioural flexibility is a core variable in the manuscript, so it should be introduced clearly to readers.

Line 68, wait, behaviour is a much broader term than behavioural flexibility but interchangeable? And how do these terms different from behaviour change (line 60)

Line 95: 'prediction 1' is not necessary because there is only one prediction

Lie 96 – trait? There are too many terms that may or may not indicating the same thing now... need clarifications here.

Methods:

1) Need more information about why some of the individuals were found at both sites? How close are Woodland and Bufferlands, can these locations be placed on figure 1?

2) The Authors make a good point of testing certain age group but the choice of age group is not entirely clearly to me as in how testing adults is related to the prediction. Say in my main study species, squirrels, the juveniles disperse the most, and they are (logically) those that show the highest flexibility and explorations, and hence, should test the younger and not the older squirrels – if they are the ones who settle in the edge of the expanding population, right? What about grackles? what is the dispersal characteristics in grackles? Providing more information about the dispersal and how it relates to the hypothesis would be helpful here.

3) Figure 2 is shown too early in the manuscript, would be better if it is placed around line 186 when talking about the experimental order. However, the figure is a bit misleading as it seems to indicate all birds went through the same order of the test, but it said the order was counterbalanced for birds. Need some clarifications in the figure captions or use A, B, C to separate the three figures in figure 2.

4) Line 124 'adults are identified... adulthood.' There is a jump between ideas, would be better if this sentence is before 'however, due to difficulties in trapping this species...'.

5) Line 127-134 'mist nets decrease the likelihood of... the fifth day instead)'... this information really should be placed close to the beginning after the sites are mentioned. By the procedure order – you got a site identified – set a new – caught the birds – it's adult! – put ring on their leg – bring in the lab to habituate the set up.

6) Innovation: would the authors think that solving more problem = higher exploration too?

7) Persistence: persistence comes out of no where, some intro why measuring this trait is important needed. The way the authors measured persistence seemed to be motivation to me... though the two terms are interchangeable.

8) Exploration. The way the authors measured exploration seems to be neophobia

9) Line 179: what is this motivation test entails?

Discussion

Line 432-435 'our measures of flexibility.... Meaningfully compared' – repeatability is an additional and informative analysis beside testing the major hypothesis. These should be placed after answering the major research question.

Line 438. Another term here that may be indicating the same or different thing of 'behavioural flexibility'...

Line 440. The authors may want to reclarify what this 'flexibility' is and how it was measured.....

Line 487. I agree with this possibility... the test may be conducted 'too late' (of grackels point of view and species characteristics). May worth adding more information the population that is in this study and suggest what would be a better idea to test this hypothesis again (what is 'long enough' line 489?).

Reviewed by anonymous reviewer 1, 24 May 2023

This is a very interesting study that examines the role of behavior, specifically behavioral flexibility and innovation, in the expansion of a species. The study aims to investigate whether individuals at the periphery of a population, in newly colonized areas, exhibit novel behaviors or rather intensify certain behaviors already present in the entire population. Overall, the findings suggest that individuals in newer populations demonstrate greater innovation but engage in less exploration. However, no differences in flexibility or persistence were observed.

Despite the interesting results, I have several general and specific considerations that could enhance the organization and comprehensibility of the study.

General considerations

On the one hand, the introduction seems too specific, I would like to see what has been studied and what behavioral variations have been found in different species. Also, I would not include the part of the great-tailed grackles in the introduction, but I would make a specific section in material and methods on the study species. In addition, the objective of the article is mentioned from the beginning, when this should only be specified at the end of the section (i.e., introduction). In the previous paragraphs, it should be mentioned which are the lack of knowledge in the subject matter.

On the other hand, I consider the material and methods section to be very complex and a bit cluttered. I consider that material and methods need to be restructured, simplifying the sections that this section is divided into and explaining in greater depth and with greater order all the experiments and the decisions taken with the individuals handled. I understand that analysis and evaluation are not easy, but I think it is important to follow the workflow.

One of my main doubts about the experiments is to know if you have considered that being a social species, learning by imitation/observation (i.e., use of social information) has an important role, which however in your experiments you are not considering it.

Also, in the behavioral flexibility analyses you include the learning rate, however I do not see that you include the option of learning or behavioral change in the rest of the experiments. Not only learning but adaptation to captivity and to the experiments themselves should be considered, and this would be easily modeled by including the number of experiments the individual has previously undergone, or the number of days in captivity (which also serves as a proxy for experience with the experiments).

Specific comments

Introduction

I believe that a more general approach to the central issue is needed: to establish the importance of behavioral flexibility in species expansion. Has it been studied in other taxa: mammals, for example? And in other species?

The more specific part of the introduction where the species is discussed as a specie of interest (lines 56-67), I would move it to a specific section of material and methods called ”Study species”.

Lines 43-44. You talk about innovation, but you do not define the concept. Also learning methods (individual or social) influence behavioral flexibility, perhaps it is important to mention it.

Lines 48-49. In these sentences you talk about what is expected, I would leave this for the end of the introduction, where the hypotheses and predictions are established. This paragraph lacks a more general perspective on the role of behavior in population dynamics.

Lines 64-67. Have neophobia levels been studied in this species, and are they lower than those of other species with other ecological requirements?

Lines 71-72. Why are you looking at variances? Specify.

Lines 75-76. Is scanning then a measure of time?

Table 1. How is the generation time calculated?

Could it be that the scan is different as a function of time or the number of experiments that have been performed on the individual? In the end, all parameters can be affected by the experience gained.

Lines 92-111. As I have already mentioned in the general comments, I think it is necessary to combine in a single paragraph the main objective of the work, then describe very concisely how it is carried out and the specific hypotheses and predictions.

Line 103-106. Try to be more direct with sentences, rather than starting them with "Perhaps in newly….." Say: "In newly established populations, individuals may need to learn about innovate new foraging techniques or find new food sources."

Line 104. The "and" is not necessary.

Line 107. "Higher variances in behavioral traits will indicate that there is…"

Methods

The methods must be written in the past tense.

What is the basis for the selection of 3 to 4 days of habituation?

The fact that it is a sociable species and that you put them individually in the aviary may not be modifying their behavior?

It is not clear how many individuals from each population are used, I recommend specifying this clearly in the text.

Were there individuals that spent different lengths of time in captivity? That could influence the outcome of the experiments.

Lines 135-137. I consider that it is not necessary to explain why field work was not carried out in other areas, although the arguments are weighty, what is important is the data obtained and analyzed.

Line 151. Regarding the flexibility protocol, do all individuals require the same number of attempts to fix the selection of a color?

Line 146. The two lines corresponding to the "Data collection stopping rule" section would be included in the previous section.

Line 171. In line with another of my previous comments: when do you consider that an individual is habituated? Are there differences between individuals in the habituation process?

Lines 176-178. I do not understand the definition of Time 1 and Time 2. Rewrite.

The section on Randomization and counterbalancing, together with that on experimental order, I would include it in the previous sections, simply adding the sentences in this regard (without titles).

Line 196. I would call the general section: statistical analyses. The other sections that follow, e.g., "flexibility analyses" or "model and simulation" should be subsections of this one.

Lines 199-201. Wouldn't this be a good opportunity to test whether the experience of participating in different experiments influences the individual's behavioral skills? I think it may be important to take into account how many experiments you have participated in and how many different types, as this can be a learning experience for the individual.

Line 242. The version of R is already specified above, remove in one of the two places.

Line 214. Can you define what the attraction score is?

Line 215. Define "rate of deviating from learned attractions".

Results

Figure 3. Did the value of learning rate change as the individual was subjected to more experiments? Perhaps this learning rate increases to a greater extent in individuals belonging to new populations.

Discussion

In general, as in the case of the introduction, I consider that the discussion lacks more general references.

Line 445. This paragraph is totally linked to the previous one. Join.

Actually, behavioral flexibility may be playing a fundamental role in both populations, as this species is so closely linked to humanized environments. Perhaps species from more stable, more natural environments (e.g., forested environments) show lower levels of flexibility? Haven't similar species living in different environments been compared (they don't have to be birds)?

Line 476. "great-tailed grackles are flexible on the reversal…" capitalize the first word.

Line 487. Is this population still expanding?

Lines 495-497. Why, if there are populations that are still establishing, were individuals from these areas not used?

LIne 503. It is not clear to me why longer settled populations are more explorative. Earlier you say that in guinea pigs no changes in this type of behavior have been seen.

Lines 524-527. I don't understand the difference between flexibility and innovation. One seems to be directly related to the other.