Dear Dr.'s Blandine Doligez, Joe Nocera, Marion Nicolaus, and Laure Cauchard,

We greatly appreciate the time you have taken to give us such useful feedback! We are very thankful for your willingness to participate in the peer review of preregistrations, and we are happy to have the opportunity to revise and resubmit.

We revised our preregistration and associated files at http://corinalogan.com/Preregistrations/gspaceuse.html, and we responded to your comments below.

Note that the version-tracked version of this preregistration is in rmarkdown at GitHub: https://github.com/corinalogan/grackles/blob/master/Files/Preregistrations/gspaceuse.Rmd. In case you want to see the history of track changes for this document at GitHub, click the previous link and then click the "History" button on the right near the top. From there, you can scroll through our comments on what was changed for each save event and, if you want to see exactly what was changed, click on the text that describes the change and it will show you the text that was replaced (in red) next to the new text (in green).

We think the revised version is much improved due to your generous feedback!

All our best,

Kelsey, Cody, Melissa, Luisa, and Corina

Does great-tailed grackle space use behavior reflect individual differences in exploration?

Kelsey B. McCune, Cody Ross, Melissa Folsom, Luisa Bergeron, Corina Logan

Recommender: Blandine Doligez

**Comment 1: I concur with the reviewers in that the question investigated here as it is stated currently relates to dispersal syndromes (i.e. consistent suite of behavioural differences between individuals) rather than behavioural flexibility (i.e. within-individual variation). Yet, as we understand, the authors plan to measure exploratory behaviour over a possibly large long of time (it is mentioned that individuals may be held in captivity up to 6 months – see below regarding this point), and this may imply a series of measures on the same individuals over time, which should allow estimating within-individual repeatability in this behaviour and thus, by contrast, flexibility (i.e. the variable part). Similarly, the movements of tagged individuals can be followed in the field over large periods of time, which should allow estimating again repeatability and by contrast flexibility. However, we currently have no information about such estimates (how many series of measurements in captivity? how to estimate within-individual variability of movement in the field, over space and time?); instead, the feeling here is that basically only one measure will be used for each behavioural compartment (exploration in captivity and space use

in the field, in particular with the use of a single measure of home range). I believe that, as presented here, the authors aim to address behavioural syndromes, but they nevertheless have the potential to explore flexibility, thus they should possibly re-consider their behavioural measures so as to make sure they estimate both the consistent and flexible part, and rewrite this estimate description and analyses parts with this in mind. I believe that assessing flexibility as stated from the start would certainly be of great interest here.**

Response 1: We only now after reading your comments realized that we forgot to remove the first, general part of the abstract that discusses our overall larger project framing around behavioral flexibility. We are so sorry for this oversight on our part and we see how it caused confusion about whether we are investigating flexibility in this preregistration. We also see how the mention of flexibility in the abstract led to confusion about whether we were talking about flexibility or behavioral syndromes (as you defined them above). We can confirm that we are addressing neither flexibility nor behavioral syndromes in the current preregistration because these pieces of the grackle project have already passed pre-study peer review at PCI Ecology (see Logan et al. 2019 and McCune et al. 2019). In the current preregistration, our aim is to investigate movement behaviors in space from an ecological perspective.

We made the following changes to the beginning of the abstract to make this clearer:

“Great-tailed grackles (*Quiscalus mexicanus*) are rapidly expanding their geographic range (@wehtje2003range) and it is generally thought that they must rely on behavioral flexibility to achieve this feat [Logan et al. 2019](http://corinalogan.com/Preregistrations/g_flexmanip.html). However, it is alternatively possible that the individuals on the range edge are more exploratory and exhibit distinct movement behaviors in space (e.g., have larger home ranges and are less predictable about which locations they visit daily), facilitating the range expansion. There is evidence for a relationship between exploratory traits and dispersal (the movement of young and/or adults into new territories; @cote2010personality), but it is still unknown whether individual differences in exploration affect daily movement patterns. This species is strongly associated with human-modified landscapes and it eats a variety of human foods (e.g., they eat our crops, at our outdoor cafes, and out of our garbage cans) in addition to foraging on insects and in the ground for natural food items (@johnson2001great). Distinct daily movement behaviors (i.e. “space use”) might facilitate range expansion above and beyond dispersal if the ability and motivation to encounter novel foods and food sources is the limiting factor in their expansion (@spiegel2017s) rather than their ability to flexibly choose among a variety of options (e.g., to keep track of which restaurants serve lunch at outdoor cafes, when the busiest times are and potentially choose among them according to preferred food type)...”

The detailed methods for how exploration is measured (in aviaries and in the wild) is in a separate preregistration (McCune et al. 2019) that we referenced in this preregistration. However, we have now added a summary of the methods here for ease of reading (Methods > Summary of methods for measuring exploration). In sum, there are two estimates per bird of

exploration in captivity, separated by approximately 6 weeks, and we will analyze in the exploration preregistration whether these are repeatable within individuals.

You are correct that we will collect space use data on grackles in the wild over a long period of time (at least 1 month). We have clarified this in:

Methods > Planned Sample: "...Therefore, now we use a leg loop harness (methods as in @rappole1991new) made from sutures (Vicryl undyed 36in sutures, item number D9389 at eSutures.com; 0.5mm diameter, absorbable **so they fall off after one to four months**)."

As a result, we are testing for repeatability of the space use variables as well. Part of our data analysis for step length and turning angle includes measures of consistency so that we will know if individuals show repeatability in these two movement behaviors. Furthermore, we will determine that home range is consistent when we conduct the asymptotic convergence analysis. If additional data points beyond 20 (@noonan2019comprehensive) do not eventually lead to asymptotic convergence in the home range estimate (see Methods > Sample size rationale), then those individuals do not have a consistent home range (@calabrese2016ctmm). We will also determine repeatability of home range estimates by comparing estimates retrieved from points taken on an individual in the breeding and non-breeding seasons.

These details are described in the preregistration in sections:

Abstract: "The first will describe individual differences in movement behavior by analyzing autocorrelation of step length (distance between two sequential observations) and turning angle for each individual over time (@pacheco2019nahua), while the second will describe individual differences in spatial preferences by analyzing the repeatability of each individual's occurrence in particular geographic locations."

Methods > Dependent variables > P1-P2, variable 4: "Spatial location preference: measured as the repeatability of grackle occurrence in a given cell of a 5 x 5m grid array across the landscape"

Analysis Plan:

"We will first verify that the GPS point locations on each bird result in asymptotic convergence as in @leo2016home. To calculate our dependent variable we will use the autocorrelated kernel density estimate method for quantifying home range size (in square meters) using the akde function in the R packages ctmm [@calabrese2016ctmm] and sf [@pebesma2018simple]. Autocorrelated kernel density estimates (AKDE) of home range size are the most accurate when data are collected close together in time and space [@noonan2019comprehensive]. We are interested in all movements by grackles, so we will not exclude any outlier relocations collected during "normal activities" [@calenge2011home]. 'Normal activities' indicate that grackles are not engaging in behaviors that would artificially skew their space use, for example mobbing a predator or the experimenter, or behavior before sunrise or after sunset when they

are at the roost. Outside of these circumstances, we will include all data to detect any space use movements."

"Secondly, we will determine whether our space use variables vary by season (breeding or non-breeding season). If season has no significant effect, all data will be included in our subsequent analyses. If there is a significant effect of season, we will run models separately for each dependent variable and each season."

"We will determine whether these parameters governing movement are stable (or variable) within individuals across days. A small variance would indicate there is low variability (high repeatability) in the daily movement behaviors of the individual."

"Moreover, we will determine whether grackles show individual differences in consistent use of habitat by overlaying a grid array across the landscape. We will then create matrices describing the number of times a grackle was observed in each cell on each day. High autocorrelation among daily matrices indicates an individual that frequents the same spatial locations across days."

Code:

Repeatability of home range in breeding and non-breeding seasons

hr_rpt = rpt(log(area) ~ (1|BirdID), grname = "BirdID", data = hr, datatype = "Gaussian", nboot = 500, npermut = 500)

summary(hr_rpt)

Again, we greatly apologize for the confusion that we caused! We hope that you find the revision and our responses below much clearer and a worthy contribution to our knowledge base.

Logan et al. 2019. http://corinalogan.com/Preregistrations/g_flexmanip.html

McCune et al. 2019. http://corinalogan.com/Preregistrations/g_exploration.html

Noonan, M. J., Tucker, M. A., Fleming, C. H., Akre, T. S., Alberts, S. C., Ali, A. H., ... & Blaum, N. (2019). A comprehensive analysis of autocorrelation and bias in home range estimation. *Ecological Monographs*, *89*(2), e01344.

Calabrese, J., Fleming, C. and Gurarie, E. (2016). Ctmm: an R package for analyzing animal relocation data as a continuous-time stochastic process. *Methods in Ecology and Evolution,* 7(9), 1124-1132.

Comment 2: In line with this, the authors should be careful about how they characterize exploration as measured by the tests in captivity and exploration as can be measured by space use behaviour in the wild. It is often assumed that prospecting movements represent exploration in the wild, but using the same term in both contexts here can be misleading about the hypotheses tested, as this gives the feeling that it is already assumed that space use is linked to exploratory – one of the very question investigated.

Response 2: Excellent point! We changed the term “exploration” as connected with space use to “use the space” (Prediction 1 alt 4), “move longer distances” (Prediction 2), and “movements” and “frequents the same spatial locations” (Analysis Plan). Therefore, the only mentions of exploration now are in connection with the exploration tests of the novel object and novel environment.

Comment 3: Indeed, space use will be largely influenced by habitat use, thus I concur with reviewers that a clear description of the habitat available for different activities (foraging, breeding, mating…) is needed here for each individual monitored. If such information is not accounted for, it will be very difficult to address the questions of interest, especially if habitat varies globally between the three populations investigated.

Response 3: We measure micro-habitat, foraging and breeding behaviors during focal follows that we conduct on each color marked individual up to 6 times per year (half in the breeding, half in the non-breeding season). These data inform a separate prepregistration on foraging behavior (http://corinalogan.com/Preregistrations/g_flexforaging.html). Therefore, we will have detailed data on the composition of the habitat in each individual’s home range and how individuals are using the habitat. Thanks to these comments, we are now including these characteristics in our analysis of spatial location preference. We added the following:

Methods > Planned sample:

To control for alternative factors contributing to space use behavior in wild grackles we will also include covariates in our models that measure energetic condition (described in this separate preregistration [Berens et al. 2019](http://corinalogan.com/Preregistrations/gcondition.html)), and habitat characteristics like human food sources and available breeding habitat (described in this separate preregistration [Logan et al. 2019](http://corinalogan.com/Preregistrations/g_flexforaging.html).

Independent variables > P1 and P1 alternatives 2-4:

5) The number of known breeding sites (shade trees, date palms, marsh vegetation (@johnson2001great) within the home range of each individual (data collected as part of [Logan et al. 2019](http://corinalogan.com/Preregistrations/g_flexforaging.html))

6) The number of human food source areas (dumpsters, cat food bowls, outdoor restaurant seating) within the home range of each individual (data collected as part of [Logan et al. 2019](http://corinalogan.com/Preregistrations/g_flexforaging.html))

7) Scaled mass index [@pieg2009new] as a measure of energetic condition

Additionally, conspecific density could affect home range size (e.g., higher population densities could result in smaller home range sizes; @flockhart2016factors, @garabedian2018relative). To account for this possibility at both the individual and population level we have incorporated additional measures for data collection and analyses.

To measure individual differences in conspecific density, we quantify the number of grackles within 10m of the focal individual during all focal follows (described in a separate preregistration [Logan et al. 2019](http://corinalogan.com/Preregistrations/g_flexforaging.html). We will take the maximum group size and include it as an independent variable in P1:

Independent variables > P1 and P1 alternatives 2-4:

8) Maximum group size observed across each individual's focal follows (data collected as part of [Logan et al. 2019](http://corinalogan.com/Preregistrations/g_flexforaging.html))

To measure the relationship between population variation in conspecific density and home range sizes, we will conduct point count surveys at each study site. We added a description of our methods as follows:

Methods > Planned Sample:

"Conspecific density has also been shown to affect home range size in other bird species (@flockhart2016factors; @garabedian2018relative). To control for the possibility that home range size may vary among our populations due to conspecific density rather than exploratory traits, we will use point count surveys to measure grackle population density. We will place 225 point count stations across the landscape encompassing each population (Tempe, AZ; Woodland, CA; Gamboa, Panama). For each study region the first central point will be randomly placed within 500 m of the center of the study area. The remaining points are placed in a 500 m grid pattern extending out from this central point. In total, the sample area will cover an area that is 7 km by 7 km. Each point will be visited once during the non-breeding season (Sep-Mar). During the survey, researchers will record all grackles seen and heard for 6 minutes."

Independent variables > P2:

4) Population density (number of grackles per square meter in each study area) (Arizona, California, Panama)

Open Materials:

[Point count protocol](https://docs.google.com/document/d/1zDuoI0v7Rv0iwrTMAZSERH7AB4v85qXzxsx5T0daNNo/edit?usp=sharing) for measuring grackle population density in the study area

Johnson, K. and Peer, B. (2001). Great-tailed Grackle. *Birds of North America,* 576, 1-28.

Comment 4: I also agree with reviewers that replicates for each zone would be needed to allow separating an individual population effect from the location within the expansion range. Alternatively, a more complete gradient of populations over the expansion range could also be used (e.g. either 2-3 replicates per population at the centre, middle and edge of the expansion range, or a gradient of 6-8 populations over the range, controlling if possible for longitude and/or latitude). I acknowledge that this may be practically difficult, but three points will likely be insufficient to fully answer the question of the link with population dynamics and expansion here.

Response 4: The sampling that we do at each site is work intensive and time consuming. Given our resources and time constraints (i.e., this is a 5-year project), logistically it is not possible to add more populations to our sample at this time. A similar study by Renee Duckworth and Alexander Badyaev (2007) documented dispersal and aggression in mountain bluebirds from 8 sites across a portion of the range. However, this study used data collected from more than 30 years of monitoring bluebird populations in the state of Montana. To our knowledge, no other researchers have measured exploration and space use in a bird species across it's range, so we believe these data are still interesting and important. However, we have updated our abstract and prediction 2 to scale back the strength of the inference that we draw from our population comparisons:

Abstract: …"Additionally, we aim to describe whether grackles from the 3 populations systematically vary in space use behavior to infer a potential relationship between space use and range expansion."

Prediction 2: …"This may relate to range expansion because some of the individuals on the leading edge of the range may use more space and move longer distances [@duckworth2007coupling]."

We have now rephrased the predictions for H2 to indicate that we will compare individuals from populations and not populations relative to their position in the geographic range (replacing language like "grackles in the edge population" with "grackles in the sample from the edge of their range"). We now only compare the distributions of parameters between populations because we won't know why populations differ. We also indicate that this would be an interesting follow up question. We updated the preregistration as follows:

Hypotheses > H2 > P2: "... However, larger-scale sampling of grackle groups across the strata of the expansion front and core range would be needed to more robustly validate the hypothesis that our cross-site differences are indicative of a broader pattern driven by the location of the expansion front."

Comment 5: I do not know what the authors plan to do with the different measures of exploration in captivity, but rather than using different measures and link them one by one to the space use measures (which will automatically increase the total number of models and thus the risk of false positive), it may be better to first check whether and how these different measures relate to each other and whether they may define an overall exploration score for a given individual. I also concur with the reviewers' comments about (1) making sure about what is measured (as mentioned, the novel environment test could also be considered as a novel object test instead) and (2) it seems a bit far-fetched to assume that space use in the wild may be related to one or the other of the exploration measures in captivity only, as presented here (as mentioned, clear predictions about why it should be so seems rather difficult to do, especially without testing directly behaviour towards novel food sources) – assessing the relations between these different measures in captivity as a first step should allow addressing this point, and I would encourage the author to include such a phase in their approach, possibly simplifying prediction 1 alternatives 2 and 3 then here.

Response 5: Good points. We failed to mention in the first submission that we will determine whether the two exploration measures correlate in a separate preregistration investigating behavioral syndromes (http://corinalogan.com/Preregistrations/g_exploration.html). This connection is now clarified in the revised current preregistration in the new section (Methods > Summary of methods for measuring exploration). To ease an understanding of the work we are conducting in this preregistration, we added the following statement to that section: "If the two exploration measures are consistent within individuals, and correlate with each other, we will choose the variable with the most data. If the two measures do not correlate, we will include both as independent variables"

1. We will be able to determine whether our captive measure of exploration indeed appears to capture exploratory traits because in the separate exploration preregistration (McCune et al. 2019), in addition to two measures of exploration (novel object and novel environment), we test grackle responses to 3 separate potentially threatening objects to measure boldness. @Reale2007integrating defines the shyness-boldness trait as an individual's reaction to any risky situation, but not novel situations. Exploration is defined as an individual's response to any new situation. By measuring responses to both familiar but risky situations, and novel situations, we can verify whether our measures of exploration are capturing exploratory or boldness traits, or something else altogether (e.g., "jingle-jangle" fallacy where multiple different labels actually apply to the same inherent trait OR one label is used but encompasses several different inherent traits @carter2013animal).

2. Thanks for pointing this out! It really helps us understand what pieces of background information are necessary to understand what we are attempting to achieve. Previous studies on birds have found a relationship between movement behavior in the wild and exploration measured in captivity using novel environment (@minderman2010novel, @dingemanse2003natal) and novel object (@mettke2002significance) tests. Therefore, we think it is reasonable to predict that behavior measured in exploration assays in captivity will relate to space use behavior in the wild. To clarify more about the background behind this study, we added:

Hypotheses > H1: "Previous studies on birds have found a relationship between movement behavior in the wild and exploration measured in captivity using novel environment (@minderman2010novel, @dingemanse2003natal) and novel object (@mettke2002significance) tests, therefore the measures we are using have the potential to be ecologically relevant. "

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

Carter, A., Feeney, W., Marshall, H., Cowlishaw, G. and Heinsohn, R. (2013). *Biological Reviews,* 88(2), 465-475.

Minderman, J., Reid, J., Hughes, M., Denny, M., Hogg, S., Evans, P. and Whittingham, M. (2010). Novel environment exploration and home range size in starlings *Sturnus vulgaris. Behavioral Ecology,* 21(6), 1321-1329.

Dingemanse, J., Both, C., Noordwijk, A., Rutten, A., and Drent, P. (2003). Natal dispersal and personalities in great tits (*Parus major*). *Proceedings of the Royal Society B: Biological Sciences,* 270(1516), 741-747.

Mettke-Hofmann, C., Winkler, H. and Leisler, B. (2002). The significance of ecological factors for exploration and neophobia in parrots. *Ethology,* 108(3), 249-272.

Comment 6: Prediction 1 alternative 4 includes the first mention that only adult individuals will be tested: why is it so? This should be specified. I also concur here with reviewers in that even among adults, there may be large differences between yearlings (and possibly 2-years old, depending on the longevity of the species) and older adults in behaviour linked to dominance, experience etc. More information should be given on this point, and also generally about the life-cycle and ecology of the species, as it is currently described nowhere in the pre-print. Yet, some of the assertions and hypotheses depend on such information. For instance Prediction 1 alternative 2 mentions different ecological constraints (habitat availability, predation, etc.) but it is currently very difficult to get an idea of the influence of these constraints. Similarly, whether grackles are territorial or live in social groups and to what extent (e.g. they may be flocking over the wintering season to find food but breed in individual territories) seems fundamental information to understand the space use patterns – and also the possible influence of captivity duration on subsequent behaviour after release (as clearly mentioned by reviewers, 6 months of captivity are likely to affect access to territory / social groups…)

Response 6: To clarify our intended sample population earlier in the preregistration, we edited the Abstract to say:

"...We will measure the space use behavior of wild, **adult** grackles using radio telemetry to find color-banded grackles and record spatial locations across time using GPS..."

Then we further clarified information about age and sex distribution of the grackles in our sample in Methods > Planned Sample:

"Note that we aim to bring only adults in to the aviaries for a cognitive test battery (that is unrelated to this preregistration) so that we are able to understand what this species is capable of, rather than testing juveniles who might still be developing their cognitive skills."

Regarding differences between yearlings and older adults, please see our Response 44 below.

Regarding potential differences between temporarily captive and never captive grackles, please see our Response 43.

Regarding adding more details about the life history and ecology of this species - thank you for suggesting this - we agree that we need to add these details. We added the following text:

Abstract: “Great-tailed grackles (*Quiscalus mexicanus*) are rapidly expanding their geographic range (@wehtje2003range) and it is generally thought that they must rely on behavioral flexibility to achieve this feat [Logan et al. 2019](http://corinalogan.com/Preregistrations/g_flexmanip.html). However, it is alternatively possible that the individuals on the range edge are more exploratory and exhibit distinct movement behaviors in space (e.g., have larger home ranges and are less predictable about which locations they visit daily), facilitating the range expansion (@spiegel2017s). There is evidence for a relationship between exploratory traits and dispersal (the movement of young and/or adults into new territories; @cote2010personality), but it is still unknown whether individual differences in exploration affect daily movement patterns. **This species is strongly associated with human-modified landscapes and it eats a variety of human foods (e.g., they eat our crops, at our outdoor cafes, and out of our garbage cans) in addition to foraging on insects and in the ground for natural food items** (@johnson2001great)...”

Hypotheses > H1: “Individual differences in measures of exploration using novel environment and novel object tasks (see separate preregistration) are related to variation in space use (measured via home range size, autocorrelation of step lengths and turning angles, or whether individuals are predictably found in the same locations) **across the breeding and non-breeding seasons … We expect space use to vary within an individual by breeding season because during the non-breeding season, this species forages in smaller groups and communally roosts in larger groups [@johnson2001great]. During the breeding season, one or more males defend a territory and females place their nests within territories to raise the young [johnson2000male]. Roaming males are also present and can obtain extra-pair copulations with females on other male's territories [johnson2000male].**”

Comment 7: The reference to Duckworth and Badyaev (2007) study in Prediction 2 is misleading here as they worked on aggressiveness and not on exploration.

Response 7: Great catch, thank you for pointing this out. We revised the text to say:

Hypotheses> H2 >Prediction 2: “the individuals on the leading edge of the range expansion may move longer distances”

Comment 8: Prediction 2 alternative 1: there could also be selection for certain phenotypes for other reasons than space use, e.g. on energy allocation trade-offs, that would in a second step, i.e. indirectly, impose constraints on space use. Maybe the authors should consider this possibility, at least theoretically (as it may be difficult to test it directly)?

Response 8: This is a good point - space use and movement behavior could be determined by energetic condition. We quantify the energetic condition (using the scaled mass index) of all grackles that we trap for a separate preregistration (http://corinalogan.com/Preregistrations/gcondition.html) that has already passed pre-study peer review at PCI Ecology. To account for the potential effect of condition on space use behavior, we now include here as an independent variable the scaled mass index. We have edited the preregistration as follows:

Methods > Planned sample:

"To control for alternative factors contributing to space use behavior in wild grackles we will also include covariates in our models that measure energetic condition (described in this separate preregistration [Berens et al. 2019](http://corinalogan.com/Preregistrations/gcondition.html)), and habitat characteristics like human food sources and available breeding habitat (described in this separate preregistration [Logan et al. 2019](http://corinalogan.com/Preregistrations/g_flexforaging.html)."

Independent variables > P1 and P1 alternatives 2-4:

7) Scaled mass index [@pieg2009new] as a measure of energetic condition

**Comment 9: Overall, the link between the two main predictions is not sufficiently clearly explained, and more specifically, I do not really understand in what prediction 2 is addressing the main question, namely the link between exploration and space use. Of course, we understand that individuals will be sampled from the different populations over the expansion range, so that ultimately, the hypothesis tested is whether individuals from different populations differ in their exploration behaviour, which influences the expansion of the species. But to me, this should be more directly and clearly stated, to make the link between the two predictions clear; in particular the sampling scheme for capturing and testing individuals in captivity should be described. We only currently know that a total of approx. 60 individuals will be tested, but we do not know the composition of this sample in terms of population of origin, as well as age, sex, etc. See also the important comments by reviewers about sample size and how it is possible to

make sure that this sample will be large enough (i.e. power analysis), especially if sex, age, population of origin etc. need to be accounted for.**

Response 9: We apologize for the confusion regarding our second hypothesis and prediction 2. Our aim in this piece is to determine whether individuals from different populations that span the current range of this species systematically vary in their space use behavior. If so, we can better understand some of the potential factors leading to the range expansion. We have updated the abstract and prediction 2 to elaborate on the link between space use and range expansion:

Abstract: …"Additionally, we aim to describe whether samples of grackles from the 3 populations systematically vary in space use behavior to infer a potential relationship between space use and range expansion."

Prediction 2: …"This may relate to range expansion because some of the individuals on the leading edge of the range may use more space and move longer distances [@duckworth2007coupling]."

We revised the Methods > Planned Sample section to make it clearer which individuals are used for the space use preregistration, how many per population, what their ages and sexes are, and what their specific backgrounds are:

"Great-tailed grackles are caught in the wild, given colored leg bands in unique combinations for individual identification, and released at their point of capture. The great-tailed grackles used in this study have one of two different backgrounds: those that do not have radio tags and those that do. First, we opportunistically track color-banded grackles that do not have a radio tag (and thus have not spent time in the aviaries) to determine whether time spent in the aviaries is related to space use behavior. When a color-banded bird is encountered, researchers track it for 20-90 minutes, recording the spatial location every 1 minute. If the bird goes out of view, researchers attempt to find it again for 15-30 minutes before moving on. Because these data are opportunistic, we do not attempt to balance for sex, but we aim to follow at least 20 non-tagged individuals in each population.

Second, those that do have radio tags (estimated 20 individuals per population) are primarily adults who we attempt to balance for sex, and who spent up to 6 months in an aviary while they participated in behavioral choice tests [Logan et al. 2019](http://corinalogan.com/Preregistrations/g_flexmanip.html) and individual differences assays, including captive measures of exploration [McCune et al. 2019](http://corinalogan.com/Preregistrations/g_exploration.html), as part of other research projects by this lab. Note that we aim to bring only adults in to the aviaries for the cognitive test battery so that we are able to understand what this species is capable of, rather than testing juveniles who might still be developing their cognitive skills. The radio tags were originally

applied to all aviary-tested birds to ensure that we could find their nest sites and track measures of reproductive success for these individuals for which we have extensive amounts of data. Now we additionally use the radio tags to collect data in this space use preregistration, which was later developed to address additional questions based on data we were already collecting. For details about the captive environment, please refer to the preregistration associated with this part of the research [McCune et al. 2019](http://corinalogan.com/Preregistrations/g_exploration.html). Before the aviary grackles are released, they are fitted with VHF radio tags..."

Regarding the sample size, please see our Response 43.

Comment 10: Again, more information is needed about the duration of captivity with regards to the experimental tests: why would 6 months be needed, and why should this vary between individuals? How many tests are performed per individual?

Response 10: Because the space use preregistration relies on data that are collected after the grackles are released back to the wild, we want to keep the focus on the wild component of the project, therefore we refer readers to the preregistrations that cover the aviary work to learn about aviary conditions and the test battery they receive in the aviary.

To answer your question though, the amount of time a grackle spends in the aviary depends on the bird and how motivated they are to participate in daily tests. The slower they participate, the longer it takes them to get through all of the tests. Additionally, in the 2018-2019 non-breeding season there was one test that took 4 months to complete, which we were not expecting. We later removed this test and they now only need about 2 months to get through their aviary tests (unless they are a slow participator).

We revised the independent variable "History" from a yes/no categorical variable to a continuous variable of how many days an individual spent in the aviaries, which should also help address questions about whether more time spent in the aviaries resulted in behavioral differences. The change we made is:

Methods > Independent variables: "History: the number of days the individual was temporarily held in the aviaries before data collection on space use began (0 indicates the grackle was only ever in the wild)"

Comment 11: Regarding the tracking protocol in the field: it is not clear to me how the authors will ensure that they can detect exploratory trips. How to make sure that there will be no bias towards already known locations, where birds may be searched for in the first place? Are there previous data about movement in this species that may help with designing search protocols? Or will experimenters search for birds until they are found (but in any case there will be a giving-up time point, and how will it be set up)? Also what is the distance of detection in this system with the tags used? This will influence the distance of approach needed for the experimenter, and thus its potential consequences on subsequent bird behaviour (e.g. will it flee if it detects the experimenter's presence?). The protocol mentions that the experimenters will not get too close, but I believe more precise information are needed on this point here (see also comments by reviewers on the comparison between tagged and color-ringed individuals, the latter requiring a relatively close approach for identification that may modify their behaviour differently compared to tagged individuals if detected from farther away; this comparison may also be influenced by the likely different resighting effort for the two categories). Also, information about tag mass with respect to bird mass should be given and justified. Overall, it will be necessary to describe the tagging and remote tracking procedure more precisely here.

Response 11: All grackles that were in the aviaries are fitted with radio tags so that we can be certain we will find them again after release. We are likely to detect at least some exploratory trips because we use radio telemetry to find individual grackles approximately 4 times per week, at different times of the day, and we follow them for 20-90 minutes. This means that when a tagged grackle moves out of visual range, we use telemetry to find it again in the new location. We search for birds until they are found and, with only 1 exception, we have so far successfully been able to track all tagged individuals for at least 4 weeks.

Detection distance of our radio tags depends on the number of buildings between the receiver and the bird. The manufacturer (Lotek) described detection distance as "several kilometers with a clear line of sight. Or you might only get a couple hundred meters if there are obstructions (buildings/hills/trees) in the way." We have frequently detected birds that are ~500m away. We increase our chances of detecting birds at large distances by searching for the signal while on top of relatively tall structures like parking garages (e.g. we were able to detect birds that were 1600m away).

Grackles in our populations are constantly surrounded by people, therefore the distance at which human presence affects their behavior is relatively small, as is seen in other urban species (e.g. @moller2009successful). For both the radio-tagged and non-radio-tagged color-banded individuals that we track, we use binoculars to confirm the color band combination and then observe grackles from at least 30m away. If at any point the grackle alarm calls while oriented towards the human observer (indicating it's behavior is directly affected by the experimenter) we immediately stop tracking that individual for the day.

We added details about our tracking procedures as follows:

Methods > Planned sample:

After release, an experimenter tracks each tagged grackle for approximately 1.5 hours on a given day, taking a GPS point every 1 minute, regardless of whether the bird moved [@cushman2005elephants]. **We aim to follow each tagged grackle at least 4 times a week to get as much data on space use as possible. Additionally we aim to track all grackles equally during morning and afternoon times. Researchers maintain a distance of at least 30m and observe the bird with binoculars so as not to influence the grackle's behavior. If the grackle alarm calls while oriented towards the researcher (indicating the researcher's presence has affected the grackle's behavior), all tracking on that individual is stopped for the day. To ensure that we capture all locations that the individual visits and not just those where they are most easily seen and followed, grackles that move out of sight during tracking are searched for with telemetry until they are found again.**

Comment 12: What is the maximum number of birds that can be expected to be monitored? The authors only mention 'as many as possible', but this will depend on the search effort in the field, and an estimate would be useful here. Also, what will be the relative effort of catching with each planned capture method? Importantly, a potential bias in individuals' exploration behaviour depending on the catching method should be a posteriori checked using the behavioural estimates in captivity.

Response 12: We will monitor at least 20 radio-tagged grackles per study site by tracking them for 20-90 minutes at least 4 times per week. We aim to obtain over 100 points on all grackles that have radio tags, and we will opportunistically obtain as many points as possible with personnel and logistical constraints on non-tagged but color-banded birds.

Currently we have more than 1 GPS point for 73 color-marked grackles. 20 of these grackles have more than 100 GPS points (14 radio-tagged grackles, 6 non-radio tagged grackles), and we expect the number of radio-tagged grackles with more than 100 points to increase to at least 28 after we release all of the grackles currently being held in the aviaries in Arizona.

Of the 97 grackles we trapped in our Arizona population, 61 individuals (63%) were caught using mist nets. This trapping method involves setting up a large, fine mesh net in the common flight corridors of grackles. We set up the net in areas (e.g. home ranges) containing several unbanded grackles who have never been caught by this trapping method. Because the fine thread that comprises the net is very difficult to see in the right conditions (e.g., in a shaded area with very little wind), grackles do not have to be particularly exploratory to get caught. The success of this method depends on how invisible the net is to the grackles. As a result, we do not believe that there is a sampling bias towards more bold or exploratory individuals that were caught with this method.

34% of the grackles in our population were caught using a bownet trapping method where grackles must approach a novel apparatus for food. In this case, it is possible we are only catching the most bold and exploratory individuals. We use this method less because it involves extensive habituation. Therefore, combining the two trapping methods makes it more likely that we are capturing individuals along a spectrum of boldness and exploration.

If we were to test exploration performance as a function of trapping method we would have a biased sample because individuals that do not participate in aviary tests are released before we do boldness and exploration tests, and replaced with new subjects. Although it is uncertain which personality traits contribute to participation in our aviary experiments, it is likely that participating grackles are relatively more bold or exploratory. We have so far not observed any difference in willingness to participate between grackles caught with the two different methods. To date we have released 7 non-participatory individuals caught with the mist net, and 4 non-participatory individuals caught with the bownet, which almost exactly reflects our trapping effort with each method (i.e. 4 birds equals 36% of the total birds released from the aviaries that did not participate in any tests).

Because we release non-participatory grackles from the aviaries (which are potentially less bold or exploratory), to prevent bias in our space use data we started applying radio tags to, and tracking these non-participatory subjects. In P1, we had already included an independent variable for time spent in the aviaries, therefore if there is a systematic difference in space use behavior between participatory and non-participatory subjects (indicating a sampling bias), then we expect to see a relationship between the time spent in the aviaries and space use.

We added more information about this in Methods > Sample Size Rationale as follows:

“Additionally, we attach radio tags to birds that do not participate in aviary tests (currently 3 individuals) and so are released early to determine whether space use behavior differs between participatory and non-participatory grackles.”

Comment 13: Besides the comment by reviewers about the planned estimate of space use, I was wondering whether grackles could exploit successive home ranges, and if yes how this may be detected here? More generally, how will the end of the sampling period for remote tracking in the field be determined? Will it depend on each individual, based on his previous behaviour, or will it be a fixed period for all individuals? If the former, how to ensure that the same biological processes will be captured for all individuals (e.g. if some individuals may regularly move over a given range while others may exploit different home ranges successively)?

Response 13: Radio-tagged grackles are intensively tracked until their radio tags fall off or the batteries die. This could take up to 4 months, but some grackles are able to remove their tags after a month. Once the tags are off, we still see these individuals and track them opportunistically. We will continue to search for non-tagged grackles until we have enough data to detect the home range of each tagged individual (approximately 20 GPS points according to simulations in @noonan2019comprehensive). We will calculate asymptotic convergence to determine when additional GPS points are no longer adding new information to the home range estimates (@leo2016home). Furthermore, successive home ranges would be detected because we will calculate the breeding and non-breeding season home range estimates separately. Lastly, we will make sure that the average tracking duration and number of points is equivalent across populations.

These details are in Methods > Data collection stopping rule:

We will stop collecting GPS location data on tagged and non-tagged birds when home ranges are fully revealed. To determine at what point home ranges have been fully revealed, we will calculate the asymptotic convergence of home range area as in @leo2016home. We will test home range asymptotic convergence for breeding season and non-breeding season movements separately (breeding season: Apr - Aug, non-breeding season: Sep - Mar).

Comment 14: Regarding independent variables: is a sex effect expected (and in this case why) or will sex be included only as a controlling factor here? The measure proposed for condition first surprised me, as I was expecting a measure of body condition such as the classical ratio mass / size (tarsus or wing length); then I understood that this is not really condition, but rather captivity experience, and I would suggest changing the term here so that there is no confusion. What about including a 'real' measure of body condition? and age? (see above and reviewer's comments). It would also be important here to first know which variables affect exploratory measures in captivity, so I believe that this should be a first step (probably included in the other protocol focused on these measures, but it would be important to mention it here as not all readers will also read the other protocols).

Response 14: Good catch! We changed the independent variable Condition to History (Methods > Independent Variables). Please see our Response 44 about age.

Regarding the inclusion of a "real" measure of body condition, please see our response 8. We are collecting body condition data for another preregistration, and we have added that here as an additional independent variable.

Finally, we are unsure whether male and female grackles will vary in their space use behaviors. However, because of the differences in breeding behavior (a male defends one territory and multiple females nest on that territory), we thought it would be important to include in our analyses in the event that there are sex differences in this species. We think the additional clarifications about the life history of this species, which includes sex differences, in Response 6 should help address this point.

Comment 15: Regarding the presentation: the list of variables in P1 alternative 1 could simply include the additional variables to avoid repetition. Among these new variables, the problem-solving performance is mentioned here but nowhere else! It would be important to explain why this variable is included here, what will be tested, etc.

Response 15: Great idea! We implemented your suggestion of removing repetition (Methods > Independent Variables > P1 alt 1). Thank you for catching the problem solving performance independent variable - we have now removed it.

Comment 16: In the analysis plan section: how do you know that no data will be missing? How do you also know that GPS locations will be recorded during 'normal activities'? (besides, what is a 'normal' activity?)

Response 16: We apologize for the confusion. We may have missing data where a few grackles had to be released from the aviaries before we completed the exploration tests (these will be noted in the protocol as exceptions in the post-study write up). We updated the sentence to say:

Analysis Plan: "We do not plan to **exclude** any data and if there are **missing** data (e.g., if a bird had to be released before collecting their data at time 2) these birds will be excluded from analyses requiring data from time 1 and 2."

We also added a clarification of what 'normal activities' are:

Analysis Plan: "'Normal activities' indicate that grackles are not engaging in behaviors that would artificially skew their space use, for example mobbing a predator or the experimenter, or behavior before sunrise or after sunset when they are at the roost. Outside of these circumstances, we will include all data to detect any space use movements."

Comment 17: This is a long list of questions and remarks, but I hope that the authors will find it useful to improve the presentation and justification of their study, and want to stress again that this is a very nice study plan that should provide many very useful insights about important questions, both on fundamental and applied points of view.

Response 17: We really, really appreciate your help in making this a stronger piece of research! We are very happy to go through all of the feedback and revise the preregistration :-)

Reviewer 1: Joe Nocera

Comment 18: This pre-registration outlines a study on whether behavioral syndromes (in this case, neophobia) are related to space-use in the wild, and by extension, to range expansion by Great-tailed Grackles. The study has been thought through very thoroughly, and the pre-study description here is well-written. I provide some thoughts and comments below that are either 1) meant to improve the delivery of the study, or 2) will be comments the authors will see again at some point. Despite my suggestions below, I think this is a great study, and the authors should move ahead with confidence!

Response 18: Thank you for your positive feedback! We look forward to addressing your comments below.

Comment 19: In a way, it is nice that Prediction 1, and its Alternative 1, are mutually exclusive results. That means that no matter what the data say, there is a story to be told. As I was reading Prediction 1, my mind automatically went to the opposite, which is what you proposed in the Alternative. Nice job.

Response 19: Thank you! We believe it is important to **a priori** consider any outcomes that might arise.

**Comment 20: Under sample size rationale, I was hoping to see a more quantitatively-based rationale. You state that 57 is the sample size, but never really state why. Moreover, you never

provide any assurance that this is statistically adequate. An a priori power analysis would be useful here.**

Response 20: We have now done a few things to address your feedback. We clarified the reason for our minimum sample size in our revision:

Methods > Sample size rationale: “We test as many birds as we can during the approximate five years of this study given that the birds are only brought into the aviaries during the non-breeding season (approximately September through March). It is time intensive to conduct the aviary test battery (2-6 months per bird at the Arizona field site), therefore we approximate that the minimum sample size for captive subjects will be 57 across the three sites (approximately 20 birds per site with half of the grackles at each site being female).”

Regarding power analyses, we added this paragraph:

Analysis Plan > Ability to detect actual effects: “To begin to understand what kinds of effect sizes we will be able to detect given our sample size limitations and our interest in decreasing noise by attempting to measure it, which increases the number of explanatory variables, we used G*Power (v.3.1, @faul2007g, @faul2009statistical) to conduct power analyses based on confidence intervals. G*Power uses pre-set drop down menus and we chose the options that were as close to our analysis methods as possible (listed in each analysis below). We realize that these power analyses are not fully aligned with our study design, however we are unaware of better options at this time. Additionally, it is difficult to run power analyses because it is unclear what kinds of effect sizes we should expect due to the lack of data on this species for these experiments.”

And we added a power analysis for each of the two hypothesis tests. Please see the power analyses results in the Analysis Plan section for H1 P1 and H2 P2. The power analyses show that we should be able to detect small effect sizes with our minimum sample size.

Comment 21: Your radio-tracking protocol requires visual confirmation before data are collected. As such, the viewer will be within eyesight of the grackle at all times. I am not as certain as you are that this will not affect the bird's behavior. Can you provide citations to support this method as having no effect?

Response 21: Thanks for pointing out that we needed to add more details here. Please see our response to comment 11, the relevant portion is also copied here for your convenience:

“Grackles in our populations are constantly surrounded by people, and so the distance at which human presence affects their behavior is relatively small, as is seen in other urban species

(@moller2009successful). Secondly, for both radio-tagged and color-banded individuals that we track, we use binoculars to confirm the color band combination and then observe grackles from at least 30m away. If at any point the grackle alarm calls while oriented towards the human observer (indicating it's behavior is directly affected by the experimenter) we immediately stop tracking that individual for the day."

Comment 22: My biggest complaint in the entire study is around your use of Minimum Convex Polygons as the metric of home range size. MCPs are notorious for overestimating home range size. And where you are collecting very fine-scale data at the GPS level, it seems odd that you would use such a coarse metric for home range. At the very least, you should consider doing one of the Kernel Density methods (especially AKDE, seeing as you are looking at autocorrelation anyway). Or Brownian Bridges. MCPs are simply not matched to the scale of the data you are collecting.

Response 22: This is a very informative comment and we appreciate your advice. We will now use autocorrelated Kernel Density estimates instead of MCPs to quantify home range size. We have updated our preregistration methods as follows:

Dependent Variables > P1 - P2:

1) Home range size (square meters): an estimate calculated using the autocorrelated-Gaussian reference function kernel density estimate (AKDE), which is the only estimate of home range that accounts for the autocorrelation due to the small time period between each of our GPS locations [@noonan2019comprehensive]. This estimate consists of the area enclosing GPS location points for an individual grackle during its normal activities.

Analysis Plan:

"… To calculate our dependent variables we will use the autocorrelated kernel density estimate method for quantifying home range size (in square meters) using the akde function in the R packages ctmm [@calabrese2016ctmm] and sf [@pebesma2018simple]. Autocorrelated kernel density estimates (AKDE) of home range size are the most accurate when data are collected close together in time and space [@noonan2019comprehensive]..."

**Comment 23: What about the habitat the birds are using? Should this not be inventoried and used as an independent variable to help explain home range size? Without it, you are assuming

that home range size is directly related to neophobia and exploration, and that all habitats are created equal. And that is certainly not true. Could more exploratory grackles be better at defending prime territories? Or worse at it, because they move around too much? I think that ignoring habitat and focusing solely on behavior is a misstep that may limit you later on.**

Response 23: We agree that habitat likely influences space use behavior. We have habitat data from a separate preregistration (Logan et al. 2019) for these home ranges and we now include them as independent variables in the analyses in the current preregistration. Since the initial submission of this preregistration, we also decided that conspecific density may affect home range sizes, therefore we are now measuring this as well. Please see our Response 3 above where we state more specifically how we are addressing these potential confounds.

Logan et al. 2019 http://corinalogan.com/Preregistrations/g_flexforaging.html

Comment 24: I wish you the best of luck in this study!

Response 24: Thank you again, we appreciate your time and thoughtful comments!

Reviewer 2: Marion Nicolaus

Comment 25: I find the abstract very misleading. The first few sentences clearly put the emphasis on 'the role of behavioural flexibility' in colonisation process and adaptation to novel environments. Yet the all study is actually about characterising individual consistency (which implies per definition a lack or limitation of flexibility) in exploration and space-use. Also as far as I can judge the evolutionary aspect will not be covered (no fitness measures will be collected). I think it would be much more relevant to place your study in the framework of dispersal syndromes where there is exist quite a large body of literature and hypotheses have been formulated (and for some empirically tested) regarding the role of individual consistent behavioural differences in population dynamics (including colonisation / range expansion).

Response 25: Regarding flexibility and individual consistency, we are so sorry for the confusion! Please see our Response 1 above for the full explanation. We greatly apologize for this oversight.

Regarding the fitness aspect, in a separate preregistration (http://corinalogan.com/Preregistrations/gdispersal.html) that has passed pre-study peer review,

we are investigating whether there are sex differences in genetic relatedness at the Arizona field site to infer potential sex biases in dispersal (dispersal at a much larger scale than what we are able to measure with radio telemetry). Therefore, this is a separate piece of the project and not one we planned on incorporating here with space use.

Regarding dispersal syndromes, to us dispersal would refer to grackles moving out of the natal area when they are young and/or as adults searching for new territories. However, this type of longer-distance movement (and tracking of young birds) is beyond the scope of this investigation. Therefore, we are not investigating the effect of consistent behavioral differences in population dynamics. We can see where this wasn't clear before so we have now clarified this:

Abstract: "...There is evidence for a relationship between exploratory traits and dispersal (the movement of young and/or adults into new territories; @cote2010personality), but it is still unknown whether individual differences in exploration relate to daily movement patterns..."

Comment 26: If the flexibility part is nonetheless included, it should be specified how it relates to individual behavioural consistency. There exist a number of studies that have suggested that behavioural types differ in level of plasticity but empirical testing is still rather limited. It could be a promising avenue to explore.

Response 26: Please see our Response 1 - sorry again for the confusion!

Comment 27: I disagree with the claim that 'behavioural flexibility is rarely directly tested at the individual level'. There exist many studies that characterise individual variation using behavioural reaction norms with the aim of quantifying individual differences in elevations (i.e. 'personality' or consistent part of the phenotype) and slopes (i.e. flexibility or plasticity). I suggest you read this literature and readjust your claim.

Response 27: Please see our Response 1 and accept our apologies for the confusion.

**Comment 28: H1: although I appreciate that the authors have thought of alternatives to their main prediction, the descriptive nature of the data will never allow to draw causal relationships

between lab-based behavioural tests and movements measured in the wild. Hence an endless list of alternative predictions could be made and I find some alternative predictions quite far-fetched. E.g. prediction 1 alternative 3: why would approach to a novel object (a pink fuzzy wire) provide information on attraction on human-provided sources of food?? If this is really one of the interpretation why not directly manipulate food types in the lab?**

Response 28: The alternative predictions we listed are the ones we think are the most plausible given the natural history of this species. We know we will not be able to infer a causal direction between our variables because they are purely correlational. However, this correlational study will provide us with information about whether our aviary measures of exploration mean anything about how these individuals use their space in the wild. This type of basic information will allow us to determine whether we can make any generalizations about space use in the wild from measures acquired in aviary tests. This question is important for increasing the relevance of the many animal personality studies conducted in captivity for conservation and ecological issues. For this reason, it is key to keep the aviary vs. wild comparison and not to conduct the whole study in the lab. These results will also help us make inferences to develop future hypotheses and interpret results from the many other studies we are conducting on this species.

To clarify the ecological relevance of Prediction 1 alternative 3, we added:

Hypotheses > Prediction 1 alternative 3: "...Much of the food grackles consume is contained within human-made packaging (e.g., grackles search in take out bags from restaurants) or enclosed in human-made containers (e.g., garbage cans), therefore they should have a reason to approach and explore new objects to determine whether they could be a new food source."

Comment 29: Prediction 1 alternative 4: if no correlations are found, how would that affect the rest of the project? Would it end there?

Response 29: If we find no correlation between aviary measures of exploration and their wild space use behavior, then we will complete the study as planned by collecting this data across the three populations to determine whether there are any population differences. If the result for each of the three populations is that there is no correlation, then we would not analyze H2, but only H1 and finish the publication process. This result (finding no correlation) would also inform our interpretation of the results from our preregistration on exploration measured in the aviary: we would be more cautious about what our exploration tests in the aviaries are actually measuring and what other variables they might relate to (preregistration available at: http://corinalogan.com/Preregistrations/g_exploration.html).

Comment 30: H2: It would be important to know the age of the different study populations, and the distance /connectivity among the 3 study sites. The reason is that whether or not differences exist in population composition may depend on the speed of the turnover of behavioural types. In blue birds (cited example), this turnover (resulting from frequency-dependent selection on aggression level) is rather slow, i.e. major population differences are found between young populations (<5 years old) and old populations (>20 years old). However in other species, an equilibrium between dispersive and less-dispersive individuals can be quickly reached (within 2 years) in which case no population differences will be found if sites are colonised for more than two years. Behavioural differences may also erode more quickly if there exists strong connectivity between populations.

Response 30: This is a great point. We added the age of the study populations and the distance among them in the new Table 1. The field sites are at least 1259 km from each other, therefore there is no connectivity among the sites in that individuals will only be observed at one field site due to them being too far apart for one grackle to travel (based on our observations of banded populations in Santa Barbara, CA and in Tempe, AZ where individuals stay in one place year round).

It is possible that the speed of turnover at the edge site, where they started breeding between 2000 and 2004 and thus has an average of 3 generations by now, has already resulted in the population consisting mostly of individuals who were not founders and therefore may not exhibit the exploration and space use behavior that may characterize founders. However, the edge site we chose was as far north as possible while operating under the constraint that we need a large enough population that exists there year round to be able to conduct such a study as this. The longest known great-tailed grackle lifespan is 12.5 years, therefore, it is possible that many of the individuals we will be able to measure will come from the 2nd generation and still may exhibit founder effects if there aren't many differences between founders and the next generation.

Comment 31: Prediction 2 alternative 2: no differences may also indicate plasticity (all individuals converge towards the same behavioural profile after dispersal). It would be important to disentangle the two options.

Response 31: We are not able to measure natal dispersal in this preregistration because we primarily only attach radio tags to adults, so we will not be able to address potential alternatives

related to development or dispersal. It is beyond the ability of our resources to include pre-dispersal juveniles in our sample for this study. If one of the adult grackles in our study dispersed, this would result in the individual moving out of our study area, which decreases the likelihood that we would be able to relocate it. Rather, we are measuring the space use behavior of grackles within their home ranges. We now clarified this point - please see Response 25. So far, all of the grackles we have tracked have stayed in the general location where they were caught (and released) except for one male.

We aim to track each tagged grackle for several months across breeding and non-breeding seasons. As a result, we will be able to determine from this long-term data whether grackles show plastic or repeatable space use behavior after the radio tags were attached. However, if grackles from the different populations have converged on one behavioral profile prior to when we attach the radio tags (e.g. during development), then we will not be able to discern from our data whether individuals converged in behavior previously, or whether they use the same amount of space independent of plasticity.

We have added this possibility to Prediction 2 alternative 2 as follows:
“There is no difference across the geographic range in the space use behavior of grackles. This suggests that, on average, all grackles may use the same amount of space, or that there is a similar distribution of individual differences in space use in each population. **Alternatively, grackles from the two populations may converge on similar space use behavior during development, however we will not be able to distinguish between these two options with our data, which is primarily from adults.**”

Comment 32:. Why are certain grackles kept for 6 months in captivity? What does their behaviour tell us once released? These birds most likely will lose their territory and/or their social group which may greatly affect their movements.

Response 32: Please see Response 43 below for an explanation of how their social and location behavior does not change after captivity. For more information about how we revised the preregistration to explain the captive tests and the duration spent in captivity per bird, please see our Responses 5, 6, 9, and 10. For more information about what their behavior tells us once they are released, please see our Response 28.

**Comment 33: Where will the caught grackles be released? In the same capturing site? Would it possible to release birds of different origins in different parts of the expansion range? This

would allow to experimentally test whether birds will attempt to return their site of origin (where they are presumably locally adapted) or stay but alter their behaviour in an adaptive manner?**

Response 33: We release the grackles at their capture site and we updated Methods > Planned Sample to say this. We currently do not have ethical permission to translocate grackles across their range. We could apply for this permission for future research, but at this point we are focused on understanding the dynamics of their natural system first.

Comment 34: It was unclear if repeated behavioural measures (in the lab) will be taken on the same caught individuals. Please specify

Response 34: Please see Response 1 - sorry for the confusion!

Comment 35: It is well established that, in birds, females and juveniles are more dispersive than adult males. How are you dealing with this heterogeneity? Is it the plan to collect a balanced sample in each site? I am wondering whether a total sample size of 57 individuals (i.e. about 20 birds per site) is enough to capture population differences given the heterogeneity among the sampled individuals (sex, age, some staying 6 months in captivity, some not).

Response 35: We are in the middle of determining which sex in this population disperses (Sevchik et al. 2019: http://corinalogan.com/Preregistrations/gdispersal.html) and we will have the results in the next couple of weeks to confirm whether this species is similar to most other bird species. We plan to collect a balanced sample at each field site: approximately 20 grackles per site that is sex balanced (note that our Arizona field site is not completely sex balanced due to the unwillingness of more females than males to participate in aviary tests). We clarified this in:

Methods > Sample size rationale: “It is time intensive to conduct the aviary test battery (2-6 months per bird at the Arizona field site), therefore we approximate that the minimum sample size for captive subjects will be 57 across the three sites **(approximately 20 birds per site with half of the grackles at each site being female)**.”

Please see our Response 20, which shows that we should have enough power to detect small differences among our sample sites, even with up to eight independent variables per model.

Comment 36: Having replicates of populations at the three parts of the expansion range would make the study much stronger. As it is now, any population differences will be confounded with population status (core, middle of expansion and range edge) which makes interpretation of results very difficult.

Response 36: Unfortunately, it is so time and resource intensive to set up each field site that we are only able to collect data from three field sites over the course of the 5 years of funding we secured for this project. We hope that these results will provide a start for future studies that can sample additional populations. Please see our Response 4 for a more complete explanation and a description of how we revised the preregistration accordingly.

Comment 37: For how long will individuals be tracked? The glue-on methods may not last as long as the harness attachment method. Pilots may be necessary to determine which method works best.

Response 37: Individuals are tracked for as long as the radio tag stays on, and the battery lasts. We currently only use the leg loop harness attachment method, which keeps the tags on for much longer than the glue on method. The harness material degrades such that it should fall off after a few months. We updated the Methods > Planned Sample to indicate that we now exclusively use the leg loop harness method, which allows individuals to be tracked for 1-4 months: “Radio tags were initially attached to the grackles by gluing them to their backs (@johnson2001great, @mong2007optimizing), however these did not stay on for very long. Therefore, now we use a leg loop harness (methods as in @rappole1991new) made from sutures and secured with crimp beads (Vicryl undyed 36in sutures, item number D9389 at eSutures.com; 0.5mm diameter, absorbable so they fall off after one to four months)”

Comment 38: P1-P2: it may be good to include age and time of year as independent variables in the models as large scale movements may be more prevalent in juveniles and in certain times of year

Response 38: Good point. Please see our response 44 for details on age in our study populations. We've now clarified our handling of data collected during the breeding (Apr - Aug) and non-breeding (Sept - Mar) season and we updated the Analysis Plan:

"...we will determine whether our space use variables vary by season (breeding or non-breeding season). If season has no significant effect, all data will be included in our subsequent analyses. If there is a significant effect of season, we will run models separately for each dependent variable and each season."

Reviewer 3: Laure Cauchard

Comment 39: Be careful with the term 'behavioural flexibility' (see Audet J-N, Lefebvre L. What's flexible in behavioral flexibility? Behavioral Ecology. 2017;28(4):943-7). If you follow the real definition of behavioural flexibility (from psychologists), an exploration task is not appropriated. Behavioural flexibility is measured through standardized tests of reversal learning, set-shifting and self-control to determine if animals are able to rapidly and efficiently adapt to different situations. An exploration task measures only the propensity of an animal to explore a novel situation.

Response 39: Please see our Response 1 above. We completely agree with you and we greatly apologize for the confusion we caused!

Comment 40: Measuring exploration: 1) Novel environment: a tent is placed in the middle of the aviary (familiar environment). Even if it is a novel environment inside the tent (which is transparent), it seems to me that this test looks like a novel object test. Moreover, the performance to this test is measured as the latency to approach to 20cm of the tent, or the closest distance. They do not enter the novel environment? So they are not exploring the novel environment. 2) Novel object: the food is still close to the novel object, you are not measuring exploration but neophobia, as you are not measuring attraction for novelty but the motivation to feed despite the novelty (the center of the aviary is not that far from the end of the aviary when the food is). I would use a different novel object for Time 2 of this test. The latency to approach to 20cm of the novel object, or the closest distance, is a good measure for neophobia.

Response 40: 1) You are right in that the novel environment is a novel object until they enter it and then it becomes a novel environment. We use the latency to approach as the response variable because we weren't sure whether the grackles would actually go inside the tent and we wanted to have the chance to have more data per bird. Indeed, a couple of grackles have entered the tent, so we do see individual variation in this test, but the variable with the most data is still the approach. We will make sure to include the point about the novel environment actually being a novel object in terms of the measure we use in the discussion in the final article of the exploration preregistration (http://corinalogan.com/Preregistrations/g_exploration.html). Thank you for helping to clarify this!

2) Thank you for your great feedback on the exploration test experimental design! Since the exploration test requires that the maintenance diet is freely available (i.e., the birds are not trying to be forced into exploring to look for food), we can only put the food as far away from the novel object as the aviary space allows. Our response variable is the latency to approach the novel object within 20cm, which is much closer to the novel object than the food is. Therefore, if the bird is on the ground solely to eat, then their presence at the food bowl would be too far away to register as coming within 20cm of the novel object/environment and it shouldn't interfere with the exploration test. We are aware of the issues around the placement of food near objects and how it can confound measures of exploration versus neophobia/boldness, therefore, for these grackles, we additionally conduct neophobia (boldness) tests where we place food next to a taxidermic predator, a rock pigeon, and a halloween cat made out of plastic and wire and we measure their latencies to eat the food (study described in McCune et al. 2019 http://corinalogan.com/Preregistrations/g_exploration.html).

Unfortunately, the exploration tests are well underway - we are almost done with sampling in Arizona - so we won't have a chance to change our experimental design to use different novel objects at Time 1 and Time 2, but this is an interesting suggestion and something we will consider when we test future populations. Many studies that measure animal personality using novel objects have similarly used the same objects in two (or more) repeated exposures and found that responses are repeatable (e.g. @mccune2018personality). If there is habituation to the object because it has become familiar after Time 1, this will likely affect all individuals similarly and we will still be able to detect overall repeatability in behavior (e.g. @reale2007integrating). A meta-analysis of a large number of studies that measured the repeatability of animal behavior found no difference in repeatability over differing numbers of observations with the same stimuli (@bell2009repeatability). One concern we have about using different objects for Time 1 and Time 2 is that it is very difficult to predict how scary a novel object will be to a grackle, therefore if we accidentally introduce objects that are much more or much less scary across the two time periods, this could obscure our ability to determine whether there are consistent individual differences with regard to these particular novel objects.

Comment 41: Measuring space use: the protocol to determine home range size from radio-tagged birds seems ok to me. What about the protocol for the other birds?

Response 41: Thank you for taking the time to review the protocol in addition to the preregistration. We determine home range size for non-tagged birds with the same method as that for radio-tagged birds, using auto-correlated kernel density estimates (AKDE; see Response 22). For both tagged and non-tagged birds, the important component is the number of GPS points on each bird. We only calculate home range size for individuals that have more than 20 GPS relocations, which simulations show is sufficient to yield the true home range area with AKDE (@noonan2019comprehensive). This is another reason that it is important for us to conduct GPS tracks on non-radio-tagged color marked birds as well. We clarified this in Methods > Data collection stopping rule by adding:

"We will stop collecting GPS location data on tagged **and non-tagged** birds when home ranges are fully revealed for data collected in both breeding and non-breeding seasons. To determine at what point home ranges have been fully revealed, we will calculate the asymptotic convergence of home range area as in @leo2016home..."

To collect GPS points on non-tagged birds, our methods for tracking are very similar to what we use for tracking tagged birds. We have updated our tracking protocol with a section detailing the methods for tracking color-banded, non-tagged grackles.

Essentially, the major difference is that with non-tagged grackles it can be difficult or impossible to find them again after they move out of sight. To make the most of our data collection time each day we must set a time cutoff for how long we search for the bird. Therefore, we attempt to relocate them for 15-30 minutes before moving on to another task or a different bird.

Comment 42: Are the birds released at the same place they have been caught?

Response 42: Yes. We updated the Methods > Planned Sample to indicate this: "Great-tailed grackles are caught in the wild and given colored leg bands in unique combinations for individual identification, and released at their point of capture..."

**Comment 43: I think it is a good idea to measure space use for grackles that stayed in the wild, to control for the possibility that grackles that were previously in the aviary have different space use behavior from non-aviary-held grackles after their release, as they can stay in

aviaries for a very long time (6 months) and they may have been forced to find a new territory, or find a new social group (it is a social species). A good sample size would be required here to rule out this possibility.**

Response 43: So far almost all of the grackles we have caught and brought into the aviaries returned to the location at which we captured them. Grackles caught in a similar location are seen again in that location in the same social groups. We now have 1.5 years of focal follow data at the Arizona field site and so far adults appear to stay in one general area. Based on these findings, it is unlikely that the aviary grackles that are released back to the wild change social groups or home ranges.

It is possible that grackles that participate in aviary tests are more bold and exploratory than conspecifics that were never in the aviaries or did not participate in aviary tests and so were released early. If this translates to differences in movement behaviors, we would be systematically biasing our data. Because of this possibility, we are comparing aviary birds that have radio tags with birds that were never in the aviaries (and who also do not have radio tags) but are findable and followable because of their color leg bands. Furthermore, we occasionally release grackles from the aviary that were not willing to participate in testing, and we attach radio tags to these individuals as well to determine whether the more participatory grackles use space differently in the wild.

Thus far, we have tracked 13 grackles that participated in aviary tests and we have more than 20 GPS relocations on 34 other grackles, 6 of which were in the aviaries but were released early because they did not participate. 3 of these 6 non-participatory grackles have radio tags. This should be a large enough sample size to be able to investigate whether there are differences in space use behavior between aviary-tested birds (n=13 so far) and birds that were never in the aviaries (n=28 so far). We also changed our “History” independent variable in the analyses to be a continuous variable, which accounts for how long a bird was in the aviary (0 days indicates birds that were only ever in the wild).

Accordingly, we revised the Methods > Sample size rationale: “We will also opportunistically collect GPS point locations on all occasions that we see any color-marked bird so we can determine whether grackles that were previously in the aviary have different space use behavior from non-aviary-held grackles after their release. We will attempt to match the sample size of aviary birds, and in our Arizona population we currently have over 20 points (the minimum number for reliably calculating home range size @noonan2019comprehensive) on 31 individuals that never had radio tags. We aim to acquire more than 20 points on at least 20 non-tagged grackles in the other two populations as well. Additionally, we attach radio tags to birds that do not participate in aviary tests (currently 3 individuals) and so are released early to determine whether space use behavior differs between participatory and non-participatory grackles.”

Comment 44: Statistics: I think you should control for age, if you can, even if you use adults only. What is the average lifetime of these birds? 1 year-old adults might differ in their space use and/or exploration level vs older birds.

Response 44: We agree that age might influence space use and exploration. We have tried to control for this by only bringing adult grackles in to the aviaries (with two exceptions where juvenile males were accidentally brought to the aviaries because their age status was not very clear) so that our question solely refers to adult behavior. The oldest grackle on record was 12.5 years (Klimkiewicz & Futcher 1987 J Field Ornithol, https://genomics.senescence.info/species/entry.php?species=Quiscalus_mexicanus). There are no average lifetime estimates available because no population has been tracked for a long enough period of time.

While Pyle (1987) suggests that it is possible to distinguish second year grackles from first year grackles, in practice we do not observe consistent differences that align with these descriptions. This is likely due to the large population differences in biometrics and other morphometrics that we are finding. Therefore, we are not able to determine the age of an adult grackle after it's first year so we cannot divide our sample into age classes other than first year and adult. We do catch first year male and female grackles, and we attempt to collect space use and exploration measures on these individuals in the wild. Thus far, we have tracked five known juveniles (2 tagged and 3 non-tagged), but because the focus is to bring primarily adults into the aviaries for testing (and these are the ones who get radio tags), the sample size for juveniles will be too small to be able to make inferences about.

Pyle, P. (1987). *Identification guide to North American passerines: a compendium of information on identifying, ageing, and sexing passerines in the hand*. Slate Creek Press.