

# Implementing a rapid geographic range expansion - the role of behavior and habitat changes

Logan CJ<sup>1\*</sup>      McCune KB<sup>2</sup>      Breen A<sup>1\*</sup>      Chen N<sup>3</sup>      Lukas D<sup>1</sup>

2020-10-06

```
#Make code wrap text so it doesn't go off the page when Knitting to PDF
library(knitr)
opts_chunk$set(tidy.opts=list(width.cutoff=60),tidy=TRUE)
```

## Affiliations:

- 1) Max Planck Institute for Evolutionary Anthropology
- 2) University of California Santa Barbara
- 3) University of Rochester

\*Corresponding author: corina\_logan@eva.mpg.de

**Cite as:** Logan CJ, McCune KB, Breen A, Chen N, Lukas D. 2020. **Implementing a rapid geographic range expansion - the role of behavior and habitat changes** (<http://corinalogan.com/Preregistrations/gxpopbehaviorhabitat.html>) In principle acceptance by *PCI Ecology* of the version on 6 Oct 2020 <https://github.com/corinalogan/grackles/blob/master/Files/Preregistrations/gxpopbehaviorhabitat.Rmd>.

**This preregistration has been pre-study peer reviewed and received an In Principle Recommendation by:**

Esther Sebastián González (2020) The role of behavior and habitat availability on species geographic expansion. *Peer Community in Ecology*, 100062. 10.24072/pci.ecology.100062

- Reviewers: Caroline Nieberding, Tim Parker, and Pizza Ka Yee Chow

## ABSTRACT

It is generally thought that behavioral flexibility, the ability to change behavior when circumstances change, plays an important role in the ability of a species to rapidly expand their geographic range (e.g., Lefebvre et al. (1997), Griffin and Guez (2014), Chow et al. (2016), Sol and Lefebvre (2000), Sol et al. (2002), Sol et al. (2005), Sol et al. (2007)). However, it is an alternative non-exclusive possibility that an increase in the amount of available habitat can also facilitate a range expansion (Hanski and Gilpin 1991; Wiens 1997). Great-tailed grackles (*Quiscalus mexicanus*) are a social, polygamous species that is rapidly expanding its geographic range (Wehtje 2003) and eats a variety of human foods in addition to foraging on insects and on the ground for other natural food items (Johnson and Peer 2001). They are behaviorally flexible (CJ Logan 2016) and highly associated with human-modified environments (Johnson and Peer 2001), thus offering an opportunity to assess the role of behavior and habitat change over the course of their expansion. We first aim to compare behavior in wild-caught grackles from three populations across their range (core of the original range, a more recent population in the middle of the northern expansion front, a very recent population on the northern edge of the expansion front) to investigate whether: 1) certain behaviors (flexibility, innovativeness, exploration, and persistence) have higher averages and variances in some populations relative to others, and 2) individuals in a more recently established population exhibit more dispersal behavior (i.e., individuals are

more likely to move away from their parents). Secondly, we aim to investigate whether habitat availability, not necessarily inherent species differences, can explain why great-tailed grackles are able to much more rapidly expand their range than their closest relative, boat-tailed grackles (*Q. major*) (Post et al. 1996; Wehtje 2003). We will examine temporal habitat changes over the past few decades using existing databases on presence/absence of both grackle species and compare habitat variables to determine whether: 3) these species use different habitats, habitat suitability and connectivity (which combined determines whether habitat is available) has increased across their range, and what proportion of suitable habitat both species occupy. Results will elucidate whether the rapid geographic range expansion of great-tailed grackles is associated with individuals differentially expressing particular behaviors and/or whether the expansion is facilitated by the alignment of their natural behaviors with an increase in suitable habitat (i.e., human-modified environments).

## INTRODUCTION

It is generally thought that behavioral flexibility, the ability to change behavior when circumstances change (see Mikhalevich et al. (2017) for theoretical background on our flexibility definition), plays an important role in the ability of a species to rapidly expand their geographic range (e.g., Lefebvre et al. (1997), Griffin and Guez (2014), Chow et al. (2016), Sol and Lefebvre (2000), Sol et al. (2002), Sol et al. (2005), Sol et al. (2007)). These ideas predict that flexibility, exploration, and innovation facilitate the expansion of individuals into completely new areas and that their role diminishes after a certain number of generations (Wright et al. 2010). In support of this, experimental studies have shown that latent abilities are primarily expressed in a time of need (e.g., Taylor et al. 2007; Bird and Emery 2009; Manrique and Call 2011; Auersperg et al. 2012; Laumer et al. 2018). Therefore, we do not expect the founding individuals who initially dispersed out of their original range to have unique behavioral characteristics that are passed on to their offspring. Instead, we expect that the actual act of continuing a range expansion relies on flexibility, exploration, innovation, and persistence, and that these behaviors are therefore expressed more on the edge of the expansion range where there have not been many generations to accumulate relevant knowledge about the environment.

It is also possible that a recent increase in the amount of available habitat can facilitate a geographic range expansion (Hanski and Gilpin 1991; Wiens 1997). A species may not need to be behaviorally flexible to move into new areas if they can continue to use the same types of habitat they are accustomed to. Human-modified environments are increasing (e.g., Liu et al. 2020; Goldewijk 2001; Wu et al. 2011), and species associated with these habitats show differences in their behavior (e.g., Ciani 1986; Chejanovski et al. 2017; Federspiel et al. 2017). These species offer an opportunity for simultaneous investigation of the roles of behavior and increased habitat availability for a rapidly increasing geographic range expansion.

To determine whether behavior is involved in a rapid geographic range expansion, direct measures of individual behavioral abilities must be collected in populations across the range of the species (see the discussion on the danger of proxies of flexibility in Logan et al. 2018). We plan to test whether behavioral flexibility and/or an increase in habitat availability play a role in the rapid geographic range expansion of great-tailed grackles (*Quiscalus mexicanus*). Great-tailed grackles are behaviorally flexible (CJ Logan 2016), rapidly expanding their geographic range (Wehtje 2003), and highly associated with human-modified environments (Johnson and Peer 2001), thus offering an opportunity to assess the role of behavior and habitat change over the course of their expansion. This social, polygamous species eats a variety of human foods in addition to foraging on insects and on the ground for other natural food items (Johnson and Peer 2001). This feature increases the ecological relevance of comparative cognition experiments that measure individual behavior abilities: grackles eat at outdoor cafes, from garbage cans, and they eat our crops. As such, they generally gain experience in the wild with approaching and opening novel objects to seek food (e.g., attempting to open a ketchup packet at an outdoor cafe, climbing into garbage cans to get french fries at the zoo, dunking sugar packets in water), which makes the tests involving human-made apparatuses ecologically relevant for this species.

We first aim to compare behavior in wild-caught great-tailed grackles from three populations across their range (core of the original range: Central America, a more recent population in the middle of the northern expansion front: Arizona, a very recent population on the northern edge of the expansion front: California). We will investigate whether certain behaviors have higher averages and variances in the edge population rela-

tive to older populations. Specifically, we will investigate behavioral flexibility measured as reversal learning of food-filled colored tube preferences (C J Logan 2016; Logan et al. 2019), innovativeness measured as the number of loci they solve to access food on a puzzlebox (Auersperg et al. 2011; Logan et al. 2019), exploration measured as the latency to approach a novel object in the absence of nearby food (Mettke-Hofmann et al. 2009; McCune KB et al. 2019), and persistence measured as the proportion of sessions they participated in during the flexibility and innovativeness experiments. We will also examine whether individuals in a recently established population (California) are more likely to move away from the location they hatched by determining whether their average relatedness (calculated using single nucleotide polymorphisms, SNPs) is lower than what we would expect if individuals move randomly (Sevchik et al. 2019).

Secondly, we aim to investigate whether habitat availability, not necessarily inherent species differences, explains why great-tailed grackles are able to much more rapidly expand their range than their closest relative, boat-tailed grackles (*Q. major*) (Post et al. 1996; Wehtje 2003). Detailed reports on the breeding ecology of these two species indicate that range expansion in boat- but not great-tailed grackles may be constrained by the availability of suitable nesting sites (Selander and Giller 1961; Wehtje 2003). Boat-tailed grackles nest primarily in coastal marshes, whereas great-tailed grackles nest in a variety of locations (e.g., palm trees, bamboo stalks, riparian vegetation, pines, oaks). However, this apparent difference in habitat breadth has yet to be rigorously quantified. Great-tailed grackles inhabit a wide variety of habitats (but not forests) at a variety of elevations (0-2134m), while remaining near water bodies, while boat-tailed grackles exist mainly in coastal areas (Selander and Giller 1961). Both species have similar foraging habits: they are generalists and forage in a variety of substrates on a variety of different food items (Selander and Giller 1961). We will use ecological niche modeling to examine temporal habitat changes over the past few decades using observation data for both grackle species from existing citizen science databases. We will compare this data with existing data on a variety of habitat variables. We identified suitable habitat variables from Selander and Giller (1961), Johnson and Peer (2001), and Post et al. (1996) (e.g., types of suitable land cover including marine coastal, wetlands, arable land, grassland, mangrove, urban), and we added additional variables relevant to our hypotheses (e.g., distance to nearest uninhabited suitable habitat patch to the north, presence/absence of water in the area). A suitable habitat map will be generated across the Americas using ecological niche models. This will allow us to determine whether the range of great-tailed grackles, but not boat-tailed grackles, might have increased because their habitat suitability and connectivity (which combined determines whether habitat is available) has increased, or whether great-tailed grackles now occupy a larger proportion of habitat that was previously available.

There could be multiple mechanisms underpinning the results we find, however our aim is to narrow down the role of changes in behavior and changes in habitats in the range expansion of great-tailed grackles. Results will elucidate whether the rapid geographic range expansion of great-tailed grackles is associated with individuals differentially expressing particular behaviors and/or whether the expansion is facilitated by the alignment of their natural behaviors with an increase in suitable habitat (i.e., human-modified environments).

## A. STATE OF THE DATA

This preregistration was written (Mar 2020) prior to collecting any data from the edge and core populations, therefore we were blind to these data. However, we were not blind to some of the data from the Arizona population: some of the relatedness data (SNPs used for Hypothesis 2 to quantify relatedness to infer whether individuals disperse away from relatives) from the middle population (Arizona) has already been analyzed for other purposes (n=57 individuals, see Sevchik et al. (2019)). Therefore, it will be considered secondary data: data that are in the process of being collected for other investigations. We have now collected blood samples from many more grackles in Arizona, therefore we will redo the analyses from the Arizona population in the analyses involved in the current preregistration. In May 2020, we completed data collection for other variables at the Arizona field site: flexibility and innovation (Logan et al. 2019), and exploration (McCune KB et al. 2019), and we will soon analyze this data, therefore it will also be considered secondary data. This preregistration was submitted in May 2020 to PCI Ecology for pre-study peer review. We received the reviews, and revised and resubmitted in Aug 2020, and it passed pre-study peer review in Oct 2020.

**Level of data blindness:** Logan and McCune collect the behavioral data (Q1) and therefore have seen this data for the Arizona population. Lukas has access to the Arizona data and has seen some of the summaries

in presentations. Chen has not seen any data.

## B. PARTITIONING THE RESULTS

We may decide to present the results from different hypotheses in separate articles. We may also decide to test these hypotheses in additional species.

## C. RESEARCH QUESTIONS

**Q1 (4 behaviors): Are there differences in behavioral traits (flexibility, innovation, exploration, and persistence) between populations across the great-tailed grackle’s geographic range? (Fig. 1 & 2).**

**Prediction 1: If behavior modifications are needed to adapt to new locations, then there will be a higher average and/or larger variance of at least some traits thought to be involved in range expansions** (behavioral flexibility: speed at reversing a previously learned color preference based on it being associated with a food reward; innovativeness: number of options solved on a puzzle box; exploration: latency to approach/touch a novel object; and persistence: proportion of trials participated in with higher numbers indicating a more persistent individual) **in the grackles sampled from the more recently established population relative to the individuals sampled in the older populations** (Table 1). Higher averages in behavioral traits indicate that each individual can exhibit more of that trait (e.g., they are more flexible/innovative/exploratory/persistent). Perhaps in newly established populations, individuals need to learn about and innovate new foraging techniques or find new food sources. Perhaps grackles require flexibility to visit these resources according to their temporal availability and the individual’s food preferences. Perhaps solving such problems requires more exploration and persistence. Higher variances in behavioral traits indicate that there is a larger diversity of individuals in the population, which means that there is a higher chance that at least some individuals in the population could innovate foraging techniques and be more flexible, exploratory, and persistent, which could be learned by conspecifics and/or future generations. *This would support the hypothesis* that changes in behavioral traits facilitate the great-tailed grackle’s geographic range expansion.

**Prediction 1 alternative 1: Human-modified environments are suitable habitat for grackles** (e.g., Selander and Giller (1961), Johnson and Peer (2001), Wehtje (2003)), **and the amount of human-modified environments has increased and is increasing** (e.g., Liu et al. (2020)). If the original behaviors exhibited by this species happen to be suited to the uniformity of human-modified landscapes (e.g., urban, agricultural, etc. environments are modified in similar ways across Central and North America), then the averages and/or variances of these traits will be similar in the grackles sampled from populations across their range (Table 1). *This supports the hypothesis* that, because this species is closely associated with human-modified environments, which may be similar across the geographic range of this species, individuals in new areas may not need to learn very much about their new environment: they can eat familiar foods and access these foods in similar ways across their range (e.g., fast food restaurant chains likely make the same food and package it in the same packaging in Central and North America, outdoor cafes and garbage cans also look the same across their range). Alternatively, it is possible that 2.9 generations at the edge site is too long after their original establishment date to detect differences in the averages and/or variances (though evidence from experimental evolution suggests that, even after 30 generations there is no change in certain behaviors when comparing domestic guinea pigs with 30 generations of wild-caught captive guinea pigs Künzl et al. (2003), whereas artificial selection can induce changes in spatial ability in as little as two generations Kotschal et al. (2013)). If the sampled individuals had already been living at this location for long enough (or for their whole lives) to have learned what they need about this particular environment (e.g., there may no longer be evidence of increased flexibility/innovativeness/exploration/persistence), there may be no reason to maintain population diversity in these traits to continue to learn about this environment. We will not be able to distinguish between these two alternatives within alternative 1 because populations closer to the northern edge of this species’ range were too small for us to establish such a field site. Both of these alternatives assume that learning is costly (e.g., Mery and Kawecki 2005), therefore individuals avoid it if they can. In the first case, individuals might not need to rely much on learning because they are attending to familiar cues across their range, therefore they only need to learn where in this new space these

cues are located. In the second case, individual learning that the founding individuals needed to rely on to move into this new space could have been lost due to potential pressure to reduce this investment as soon as possible after moving to a new location.

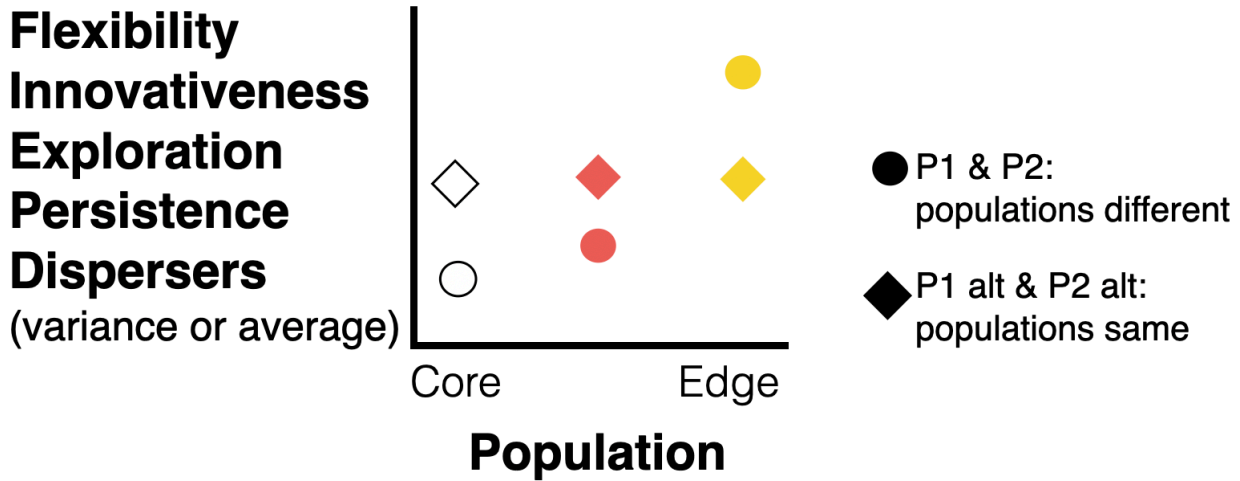


Figure 1: **Figure 1.** What is the role of behavior in a rapid range expansion? The great-tailed grackle study sites are indicated by the colored circles: edge (yellow; California), middle (red; Arizona), and core (white; Central America) and correspond with those in Figure 3.

**Q2 (dispersal behavior): Are there differences in dispersal behavior across the great-tailed grackle's geographic range? (Fig. 1, Table 1)**

**Prediction 2:** We predict **more dispersal at the edge**: a higher proportion of individuals, particularly females, which is the sex that appears to be philopatric in the middle of the range expansion (Sevchik et al. 2019), disperse in a more recently established population and, accordingly, fewer individuals are closely related to each other. *This would support the hypothesis* that changes in dispersal behavior are involved in the great-tailed grackle's geographic range expansion.

**Prediction 2 alternative 1:** We predict that the **proportion of individuals dispersing is not related to when the population established** at a particular site and, accordingly, the average relatedness is similar across populations. *This supports the hypothesis* that the original dispersal behavior was already well adapted in this species to facilitate a range expansion.

**Table 1.** Population characteristics for each of the three field sites in Q1 and Q2. The number of generations at a site is based on a generation length of 5.6 years for this species (International (2018)) and on the first year in which this species was reported to breed at the location (Wehtje (2003) for Arizona, Steve Hampton's pers. comm. reported in Pandolfino et al. (2009) for Woodland, California) (note: this species starts breeding at age 1). The first confirmed nest sighting in Woodland, California was reported in the Yolo Audubon Society's newsletter *The Burrowing Owl* (July 2004), which Steve Hampton shared with Logan. For Central America, there is no data on the first year in which they started breeding because this species originates in this region, therefore we used the age of the species: 800,000 years (Johnson and Cicero (2004)).

```
d <- read.csv(url("https://raw.githubusercontent.com/corinalogan/grackles/master/Files/Preregistrations",
header = F, sep = ",", stringsAsFactors = F))

colnames(d) <- c("Site", "Range position", "Breeding since",
"Number of years breeding", "Average number of generations",
"Citation")
```

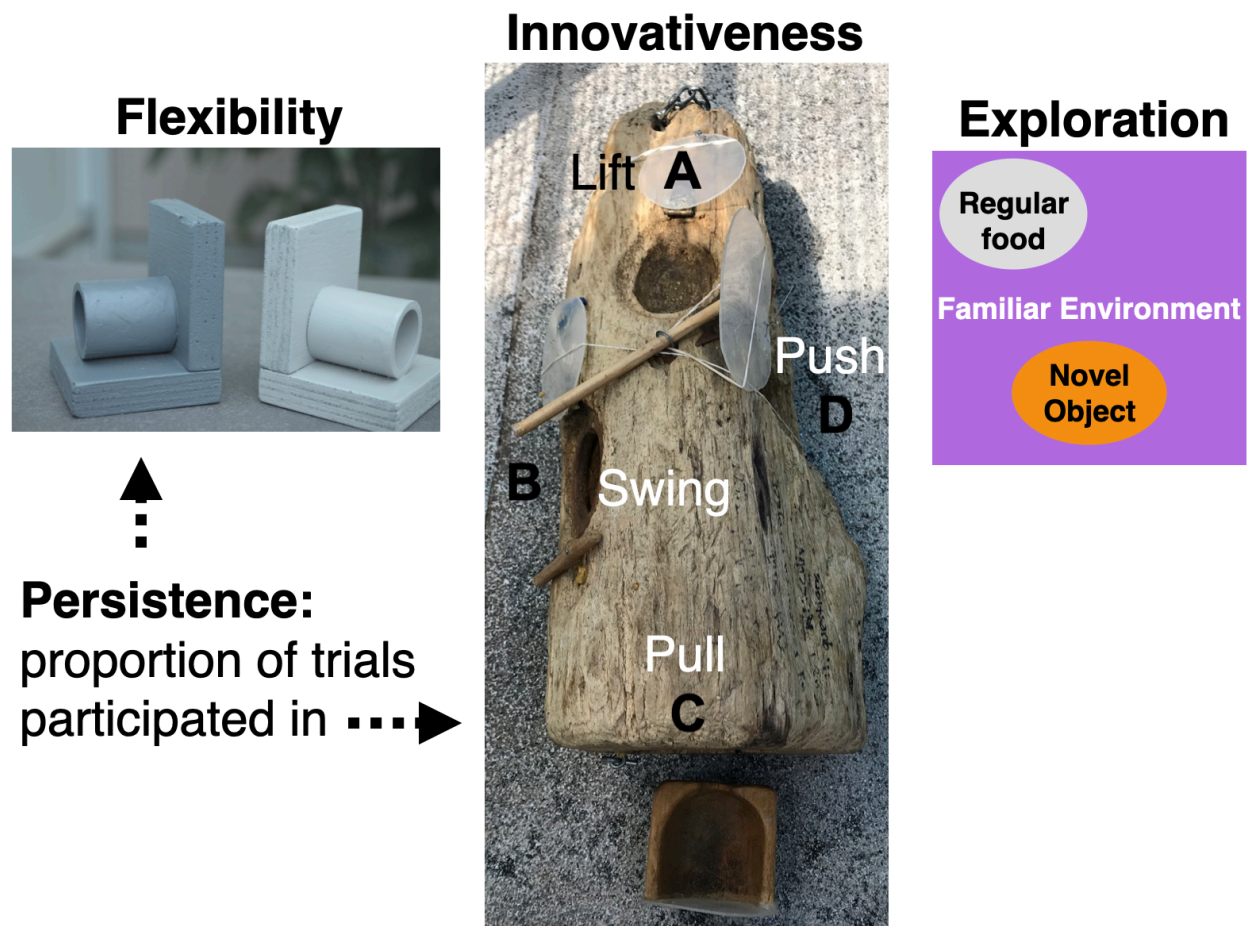


Figure 2: **Figure 2.** Experimental protocol. Great-tailed grackles from the core, middle, and edge populations will be tested for their: (top left) flexibility (number of trials to reverse a previously learned color tube-food association); (middle) innovativeness (number of options [lift, swing, pull, push] solved to obtain food from within a multi-access log); (bottom left) persistence (proportion of trials participated in during flexibility and innovativeness tests); and (far right) exploration (latency to approach/touch a novel object).

```
library(kableExtra)
knitr::kable(d) %>% kable_styling(full_width = T, position = "left")
```

Site	Range position	Breeding since	Number of years breeding	Average number of generations	Citation
Central America	Core	Unknown	800000	142857.1	Johnson & Cicero 2004
Tempe, Arizona	Middle of expansion	1936	66	11.8	Wehtje 2003
Woodland, California	Northern edge	2004	16	2.9	Burrowing Owl July 2004, Pandolfino et al. 2009

**Q3 (habitat GTGR & BTGR):** Are there differences in the availability of habitat between great-tailed grackles (GTGR) and boat-tailed grackles (BTGR)? (Fig. 3; Wehtje (2003), Selander and Giller (1961))

**Prediction 3:** GTGR and BTGR use different habitats, and the habitat of GTGR, but not that of BTGR, has increased in suitability and connectivity over the past few decades. *This supports the hypothesis* that the availability of habitat, not inherent species differences, explains why great-tailed grackles (GTGR) are able to much more rapidly expand their range than boat-tailed grackles (BTGR). Species distribution models generally do not account for additional factors such as dispersal limitations when estimating suitable habitat. Therefore, we plan to conduct two separate analyses: one to examine suitable habitat, and another to examine connected habitat.

**Prediction 4:** Over the past few decades, GTGR has increased the habitat breadth that they can occupy, whereas BTGR continues to use the same limited habitat types. *This supports the hypothesis* that changes in behavioral traits facilitate the great-tailed grackle's geographic range expansion.

**Prediction 5:** Some inherent trait allows GTGR to expand even though both species have unused habitat available to them. *This supports the hypothesis* that the original behavior of GTGR was already well adapted to facilitate a range expansion.

## D. METHODS

### Planned Sample (Q1 & Q2)

Great-tailed grackles are caught in the wild in Woodland, California and at a site to be determined in Central America. We aim to bring adult grackles, rather than juveniles, temporarily into the aviaries for behavioral choice tests to avoid the potential confound of variation in cognitive development due to age, as well as potential variation in fine motor-skill development (e.g., holding/grasping objects; early-life experience plays a role in the development of both of these behaviors; e.g., Collias and Collias (1964), Rutz et al. (2016)) with variation in our target variables of interest. Adults will be identified from their eye color, which changes from brown to yellow upon reaching adulthood (Johnson and Peer 2001). We apply colored leg bands in unique combinations for individual identification. Some individuals (~20) are brought temporarily into aviaries for behavioral choice tests, and then are released back to the wild at their point of capture. We catch grackles with a variety of methods (e.g., walk-in traps, mist nets, bow nets), some of which decrease the likelihood of a selection bias for exploratory and bold individuals because grackles cannot see the traps (i.e., mist nets). Grackles are individually housed in an aviary (each 244cm long by 122cm wide by 213cm tall) for a maximum of six months where they have *ad lib* access to water at all times and are fed Mazuri Small Bird maintenance diet *ad lib* during non-testing hours (minimum 20h per day), and various other food items (e.g., peanuts, bread) during testing (up to 4h per day per bird). Individuals are given three to four days to habituate to the aviaries and then their test battery begins on the fourth or fifth day (birds are usually tested six days per week, therefore if their fourth day occurs on a day off, they are tested on the fifth day instead).



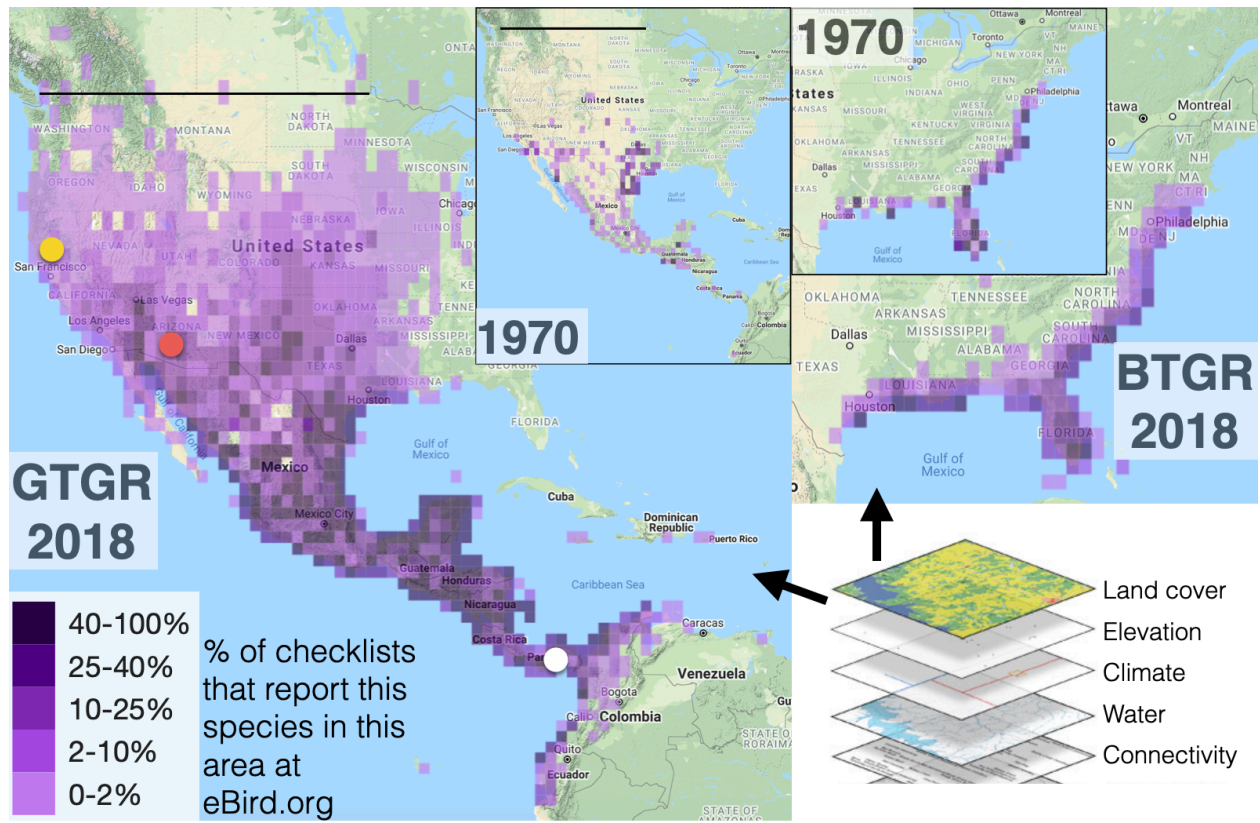


Figure 3: **Figure 3.** Ecological niche modeling of the presence/absence of great-tailed grackles (GTGR; left), which are rapidly expanding their geographic range (see inset for range change since 1970), and boat-tailed grackles (BTGR; right), which are not (see inset for range change since 1970). The four presence/absence map layers (two per species times two time periods) will be combined with habitat layers for each time period (land cover, elevation, climate, water, and connectivity) to determine habitat availability. The GTGR study sites are marked with colored circles: edge (yellow; California), middle (red; Arizona), and core (white; Central America). Map credit: eBird.org.



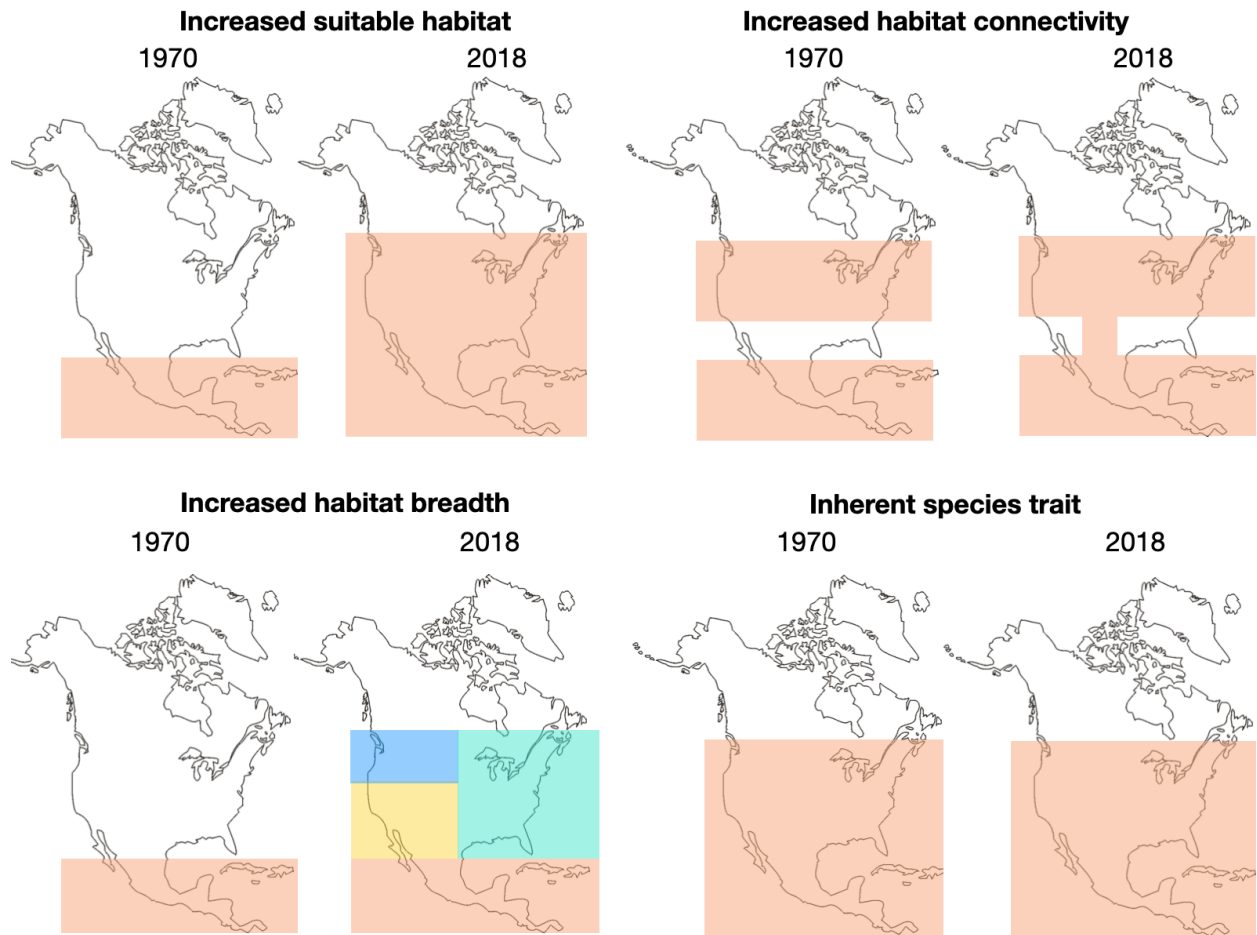


Figure 4: **Figure 4.** Ecological niche modeling will allow us to answer our research questions about the role of recent habitat changes (1970 through 2018) for the rapid range expansion of GTGR. We will assess whether the expansion of GTGR was facilitated by changes in suitable habitat (P3), habitat connectivity (P3), and habitat breadth (P4), or whether differences might be due to an inherent species trait(s) (P5). We will perform the same analyses on BTGR to determine whether they have not rapidly expanded due to a lack of suitable habitat.

While the above is our ideal plan, due to restrictions around COVID-19, it may not be possible for us to accomplish all of our goals within our current funding period. We think it will be possible to collect data at one more site (which would be the second of three planned sites) and we will attempt to also include a third field site.

### Sample size rationale (Q1 & Q2)

We test as many birds as we can during the approximately one year we spend at each site 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 at each site will be 20 grackles with the aim that half of the grackles tested at each site are female.

### Data collection stopping rule (Q1 & Q2)

We will stop collecting data on wild-caught grackles in Q1 and Q2 (data for Q3 are collected from the literature) once we have completed one year at each of the California and Central America sites (likely complete in summer 2022), which coincides with the period in which we currently have funding (until early 2023). If we are not able to collect data at a third site, we will attempt to collect more data during a second year at the second site (Woodland, CA).

### Protocols and open materials

- Experimental protocols for Q1 are online here.
- **Flexibility** protocol (from Logan et al. (2019)) using reversal learning with color tubes. Grackles are first habituated to a yellow tube and trained to search for hidden food. A light gray tube and a dark gray tube are placed on the table or floor: one color always contains a food reward (not visible by the bird) while the other color never contains a reward. The bird is allowed to choose one tube per trial. An individual is considered to have a preference if it chose the rewarded option at least 85% of the time (17/20 correct) in the most recent 20 trials (with a minimum of 8 or 9 correct choices out of 10 on the two most recent sets of 10 trials). We use a sliding window in 1-trial increments to calculate whether they passed after their first 20 trials. Once a bird learns to prefer one color, the contingency is reversed: food is always in the other color and never in the previously rewarded color. The flexibility measure is how many trials it takes them to reverse their color preference using the same passing criterion.
- **Innovativeness** protocol (from Logan et al. (2019) and based on the experimental design by Auersperg et al. (2011)) using a multi-access log. Grackles are first habituated to the log apparatus with all of the doors locked open and food inside each locus. After habituation, the log, which has four ways of accessing food (pull drawer, push door, lift door up, swing door out), is placed on the ground and grackles are allowed to attempt to solve or successfully solve one option per trial. Once a bird has successfully solved an option three times, it becomes non-functional (the door is locked open and there is no food at that locus). The experiment ends when all four loci become non-functional, if a bird does not come to the ground within 10 min in three consecutive test sessions, or if a bird does not obtain the food within 10 min (or 15 min if the bird was on the ground at 10 min) in three consecutive test sessions.
- **Persistence** is measured as the proportion of trials participated in during the flexibility and innovativeness experiments (after habituation, thus it is not confounded with neophobia). The higher the number, the more persistent they are. This measure indicates that those birds who do not participate as often are less persistent in terms of their persistence with engaging with the task. We generally offer a grackle the chance to participate in a trial for 5 min. If they don't participate within that time, we record -1 in the data sheet, the apparatus is removed and the trial is re-attempted later.
- **Dispersal:** DNA is collected from the grackles, processed, and analyzed for pairwise relatedness using ddRADseq and Stacks as in Sevchik et al. (2019) (protocol).

- **Habitat:** We will conduct ecological niche modeling to investigate grackle presence as it overlaps with available habitat across their range. Grackles will be considered as present or absent in a particular geographic area based on sightings reported at eBird.org. We identified suitable habitat variables from Selander and Giller (1961), Johnson and Peer (2001), and Post et al. (1996) (e.g., types of suitable land cover including wetlands, marine coastal, arable land, grassland, mangrove, urban), and we added additional variables relevant to our hypotheses (e.g., distance to nearest uninhabited suitable habitat patch to the north, presence/absence of water in the area). A suitable habitat map will be generated across the Americas using ecological niche models.

## Open data (Q1 & Q2)

When the study is complete, the data will be published in the Knowledge Network for Biocomplexity's data repository.

## Randomization and counterbalancing (Q1 & Q2)

**Experimental order:** The order of experiments, reversal learning or multiaccess log, will be counterbalanced across birds within a site.

**Reversal learning:** The first rewarded color in reversal learning is counterbalanced across birds at each site. The rewarded option is pseudorandomized for side (and the option on the left is always placed first). Pseudorandomization consists of alternating location for the first two trials of a session and then keeping the same color on the same side for at most two consecutive trials thereafter. A list of all 88 unique trial sequences for a 10-trial session, following the pseudorandomization rules, will be generated in advance for experimenters to use during testing (e.g., a randomized trial sequence might look like: LRLRLRLRLR, where L and R refer to the location, left or right, of the rewarded tube). Randomized trial sequences will be assigned randomly to any given 10-trial session using a random number generator (random.org) to generate a number from 1-88.

## Blinding during analysis

Blinding is usually not involved in the final analyses because the experimenters collect the data (and therefore have seen some form of it) and run the analyses. Hypothesis- and data-blind video coders are recruited to conduct interobserver reliability of 20% of the videos for each experiment.

## E. ANALYSIS PLAN

We use **simulations** and design customized **models** to determine what sample sizes allow us to detect differences between sites (see chapter 5.3 in Bolker (2008) for why simulations perform more powerful power analyses). We do not plan to **exclude** any data and if there are **missing** data (e.g. if a bird participated in one of the two experiments, then it will only be included in those analyses for which it has data). Analyses will be conducted in R (current version 3.6.1; R Core Team (2017)) and Stan (version 2.18, Carpenter et al. (2017)).

## Interobserver reliability of dependent variables

To determine whether experimenters coded the dependent variables in a repeatable way, hypothesis-blind video coders will first be trained in video coding the dependent variables (reversal learning and multiaccess log: whether the bird made the correct choice or not; exploration: latency to approach), requiring a Cohen's unweighted kappa (reversal and multiaccess categorical variables) or an intra-class correlation coefficient (ICC; exploration continuous variable) of 0.90 or above to pass training. This threshold indicates that the two coders (the experimenter and the video coder) agree with each other to a high degree (kappa: Landis and Koch (1977), using the psych package in R Revelle (2017); ICC: Hutcheon et al. (2010), using the irr package in R: Gamer et al. (2012)). After passing training, the video coders will code 20% of the videos for each experiment and the kappa and ICC will be calculated to determine how objective and repeatable scoring was for each variable, while noting that the experimenter has the advantage over the video coder

337 because watching the videos is not as clear as watching the bird participate in the trial from the aisle of the  
338 aviaries. The unweighted kappa is used when analyzing a categorical variable where the distances between  
339 the numbers are meaningless (0=incorrect choice, 1=correct choice, -1=did not participate), and the ICC is  
340 used for continuous variables where distances are meaningful (e.g., if coders disagree by a difference of 2 s  
341 rather than 5 s, this is important to account for).

```
library(irr) #ICC package

#### FLEXIBILITY (reversal learning)

# did video coder pass interobserver reliability training?
data <- read.csv("", header = TRUE, sep = ",", stringsAsFactors = FALSE)
head(data) #Check to make sure it looks right
# Note: c(3,5) is telling R to look at columns 2 and 3 and
# compare them. Double check this:
data[, 3] #coder 1 (live coder)
data[, 5] #coder 2 (video coder)
cohen.kappa(data[, c(3, 5)], w = NULL, n.obs = NULL, alpha = 0.05,
  levels = NULL)

# video coder score for 20% of videos =
data <- read.csv("", header = TRUE, sep = ",", stringsAsFactors = FALSE)
head(data) #Check to make sure it looks right
# Note: c(3,5) is telling R to look at columns 2 and 3 and
# compare them. Double check this:
data[, 3] #coder 1 (live coder)
data[, 5] #coder 2 (video coder)
cohen.kappa(data[, c(3, 5)], w = NULL, n.obs = NULL, alpha = 0.05,
  levels = NULL)

#### INNOVATIVENESS (multiaccess log)

# did video coder pass interobserver reliability training?
data <- read.csv("", header = TRUE, sep = ",", stringsAsFactors = FALSE)
head(data) #Check to make sure it looks right
# Note: c(3,5) is telling R to look at columns 2 and 3 and
# compare them. Double check this:
data[, 3] #coder 1 (live coder)
data[, 5] #coder 2 (video coder)
cohen.kappa(data[, c(3, 5)], w = NULL, n.obs = NULL, alpha = 0.05,
  levels = NULL)

# video coder score for 20% of videos =
data <- read.csv("", header = TRUE, sep = ",", stringsAsFactors = FALSE)
head(data) #Check to make sure it looks right
# Note: c(3,5) is telling R to look at columns 2 and 3 and
# compare them. Double check this:
data[, 3] #coder 1 (live coder)
data[, 5] #coder 2 (video coder)
cohen.kappa(data[, c(3, 5)], w = NULL, n.obs = NULL, alpha = 0.05,
  levels = NULL)
```

```
#### EXPLORATION (latency to approach a novel object)

# did video coder pass interobserver reliability training?
data <- read.csv("", header = TRUE, sep = ",", stringsAsFactors = FALSE)
head(data) #Check to make sure it looks right
# Note: c(3,5) is telling R to look at columns 2 and 3 and
# compare them. Double check this:
data[, 3] #coder 1 (live coder)
data[, 5] #coder 2 (video coder)
icc(data[, c(3, 5)], model = "oneway", type = "consistency",
    unit = "single", conf.level = 0.95)

# video coder score for 20% of videos =
data <- read.csv("", header = TRUE, sep = ",", stringsAsFactors = FALSE)
head(data) #Check to make sure it looks right
# Note: c(3,5) is telling R to look at columns 2 and 3 and
# compare them. Double check this:
data[, 3] #coder 1 (live coder)
data[, 5] #coder 2 (video coder)
icc(data[, c(3, 5)], model = "oneway", type = "consistency",
    unit = "single", conf.level = 0.95)
```

## Q1: behavior across the range

### Response variables

- 1) Flexibility: number of trials to reverse a color preference.
- 2) Innovativeness: total number of loci solved on the multiaccess log (maximum=4)
- 3) Exploration: Latency to approach up to 20cm of an object (novel or familiar, that does not contain food) in a familiar environment (that contains maintenance diet away from the object) - OR - closest approach distance to the object (choose the variable with the most data for the analysis).
- 4) Persistence: proportion of trials participated in during the flexibility and innovativeness experiments

One model will be run for each response variable

### Explanatory variable

There is no explanatory variable: we will conduct pairwise comparisons across sites as described in the next section.

## Hypothesis-specific mathematical model

Following procedures in McElreath (2016), we constructed a **hypothesis-appropriate mathematical model** for each of the response variables that examines differences in the response variable between sites. These models take the form of:

$$y \sim \alpha[\text{site}]$$

y is the response variable (flexibility, innovation, exploration, or persistence). There will be one intercept,  $\alpha$ , per site and we will estimate the site's average and standard deviation of the response variable.

We formulated these models in a Bayesian framework. We determined the priors for each model by performing prior predictive simulations based on ranges of values from the literature to check that the models are covering the likely range of results.

We will then perform pairwise contrasts to determine at what point we will be able to detect differences between sites by manipulating sample size, and  $\alpha$  means and standard deviations. Before running the simulations, we decided that a model would detect an effect if 89% of the difference between two sites is on the same side of zero (following McElreath (2016)). We are using a Bayesian approach, therefore comparisons are based on samples from the posterior distribution. We will draw 10,000 samples from the posterior distribution, where each sample will have an estimated mean for each population. For the first contrast, within each sample, we subtract the estimated mean of the edge population from the estimated mean of the core population. For the second contrast, we subtract the estimated mean of the edge population from the estimated mean of the middle population. For the third contrast, we subtract the estimated mean of the middle population from the estimated mean of the core population. We will now have samples of differences between all of the pairs of sites, which we can use to assess whether any site is systematically larger or smaller than the others. We will determine whether this is the case by estimating what percentage of each sample of differences is either larger or smaller than zero. For the first contrast, if 89% of the differences are larger than zero, then the core population has a larger mean. If 89% of the differences are smaller than zero, then the edge population has a larger mean.

We ran these analyses in R (current version 3.6.1; R Core Team (2017)) and used the following R packages: rethinking (McElreath (2020)), rstan (Stan Development Team (2020)), and Rcpp (Eddelbuettel and François (2011)).

## Flexibility analysis

### Model and simulation

Expected values for reversal learning using color tubes (mean, standard deviation, and range of number of trials to reverse a color preference) were based on previously published data on great-tailed grackles (C J Logan (2016)). This data indicates that the average number of trials to reverse a preference is 91 and the standard deviation is 21 ( $n=7$  grackles). The  $\sigma$  prior is set to produce only positive values that encompass the range of values shown by the Santa Barbara grackles (reversing in 70-130 trials, 130 trials-91 mean is about 40 trials).

$y \sim \alpha[\text{site}]$  *[the model]*

$\alpha[\text{site}] \sim \text{Normal}(\mu, \sigma)$   *[ $\alpha_1$  prior]*

$\mu \sim \text{Normal}(91, 21)$   *[ $\mu$  prior]*

$\sigma \sim \text{Uniform}(0, 40)$   *[ $\sigma$  prior]*

```
library(rethinking)
library(rstan)

# Plot priors to make sure the prior behaves as expected
# (p.82 Rethinking)
curve(dnorm(x, 91, 21), from = 20, to = 200) #mu result as expected
curve(dunif(x, 0, 50), from = -10, to = 60) #sigma result as expected

# Prior predictive simulation (p.82 Rethinking)
sample_mu <- rnorm(10000, 91, 21)
sample_sigma <- runif(10000, 0, 40)
prior_t <- rnorm(10000, sample_mu, sample_sigma)
dens(prior_t) #Result: behaving as expected

### SIMULATION: Manipulate mu mean, mu sd, and n (grackles)
### Simulate data for the model
n <- 60 #number of grackles in total across all sites (~20/site; AZ=20)
sites <- matrix(nrow = n, ncol = 4) #store data here
```



```

colnames(sites) <- c("trials", "mu", "sigma", "site")

for (i in 1:(n/3)) {
  mu1 <- rnorm(1, 91, 22) #mu for site 1
  sigma1 <- runif(1, 0, 40) #sigma for site 1
  trials <- rnorm(1, mu1, sigma1) #distribution for trials

  sites[i, 1] <- trials
  sites[i, 2] <- mu1
  sites[i, 3] <- sigma1
  sites[i, 4] <- 1 #site
}

for (i in ((n/3) + 1):((n/3) + (n/3))) {
  mu2 <- rnorm(1, 120, 22) #mu for site 2
  sigma2 <- runif(1, 0, 40) #sigma for site 2
  trials <- rnorm(1, mu2, sigma2) #distribution for trials

  sites[i, 1] <- trials
  sites[i, 2] <- mu2
  sites[i, 3] <- sigma2
  sites[i, 4] <- 2 #site
}

for (i in ((n/3) + (n/3) + 1):n) {
  mu3 <- rnorm(1, 77, 22) #mu for site 3
  sigma3 <- runif(1, 0, 40) #sigma for site 3
  trials <- rnorm(1, mu3, sigma3) #distribution for trials

  sites[i, 1] <- trials
  sites[i, 2] <- mu3
  sites[i, 3] <- sigma3
  sites[i, 4] <- 3 #site
}

### THE MODEL
dat <- list(trials = as.numeric(sites[, 1]), site = as.integer(sites[,
4])) #This sets up the data sheet with the simulated data that the model will run on

s1 <- ulam(alist(trials ~ dnorm(mu[site], sigma), mu[site] ~
  dnorm(91, 21), sigma ~ dunif(0, 40)), data = dat, log_lik = TRUE,
  messages = FALSE) #This is the model

# Plot it
labels <- paste("mu[", 1:3, "]:", levels(dat$site), sep = "")
plot(precis(s1, depth = 2, pars = "mu"), labels = labels, xlab = "expected trials to reverse") #contra

# Summary data
precis(s1, depth = 2) #depth=2 means to show any vector parameters (p.156)

# Compute pairwise contrasts Comparing sites 1 and 2
post1 <- extract.samples(s1)
post1$diff_12 <- post1$mu[, 1] - post1$mu[, 2] #subtract mu for site 2 from mu for site 1. col1=site 1

```

```

precis(post1, depth = 2) #look at the last row: diff_12: this is the expected difference between the d
# hist(post1$diff_12) #another way of visualizing it

# Comparing sites 1 and 3
post2 <- extract.samples(s1)
post2$diff_13 <- post2$mu[, 1] - post2$mu[, 3]
precis(post2, depth = 2)

# Comparing sites 2 and 3
post3 <- extract.samples(s1)
post3$diff_23 <- post3$mu[, 2] - post3$mu[, 3]
precis(post3, depth = 2)

### To make the model work, need to set up a few things...
### (this took me a few days because at every stage there is an
### error message and it isn't clear what the problem is or
### what to do next)

### Update R install installr package updateR(TRUE) #didn't
### work bc it can't find my R folder on my computer
### updateR(fast = FALSE, browse_news, install_R,
### copy_packages, copy_Rprofile.site, keep_old_packages,
### update_packages, start_new_R, quit_R, print_R_versions =
### TRUE, GUI = TRUE, to_checkMD5sums = FALSE,
### keep_install_file = FALSE, download_dir =
### '/Users/corina/Library/R', silent = FALSE, setInternet2 =
### TRUE, cran_mirror = 'https://cran.rstudio.com/') #didn't
### work bc it can't find my R folder on my computer manually
### updated R and reinstalled packages

### Install rethinking install package devtools
### devtools::install_github('rmcelreath/rethinking',ref='Experimental')

### Install xcode (first download the app from the app store)
### In terminal, type: 'xcode-select --install'

### Get STAN working https://github.com/rmcelreath/rethinking
### Install C++ toolchain and configuration
### https://github.com/stan-dev/rstan/wiki/Installing-RStan-from-source-on-a-Mac
### install.packages('Rcpp', repos =
### 'https://rcppcore.github.io/drat') Then install rstan per
### instructions at the previous link

### Now we're ready to go!

#### NOW keep means the same across sites and change the mu
#### standard deviations, reduce the sigma variance so we expect

```

```

#### less measurement error (sigma=10) otherwise the large
#### expectation of measurement error (previously sigma=40)
#### overshadows any actual differences attributable to SD
#### differences across individuals within a site

## Prior predictive simulation (p.82 Rethinking) because we
## are changing sigma from 40 to 10 sigma 40
sample_mu <- rnorm(10000, 91, 21)
sample_sigma <- runif(10000, 0, 40)
prior_t <- rnorm(10000, sample_mu, sample_sigma)
dens(prior_t) #Result: behaving as expected

# sigma 10
sample_mu10 <- rnorm(10000, 91, 21)
sample_sigma10 <- runif(10000, 0, 10)
prior_t10 <- rnorm(10000, sample_mu10, sample_sigma10)
dens(prior_t10) #Result: similar to previous, but distribution is a little narrower (still in the right

### SIMULATION: Manipulate mu sd Simulate data for the model
n <- 45 #number of grackles in total across all sites (~20/site; AZ=20)
sites <- matrix(nrow = n, ncol = 4) #store data here
colnames(sites) <- c("trials", "mu", "sigma", "site")

for (i in 1:(n/3)) {
  mu1 <- rnorm(1, 91, 15) #mu for site 1 stays the same. Vary SD
  sigma1 <- runif(1, 0, 10) #sigma for site 1
  trials <- rnorm(1, mu1, sigma1) #distribution for trials

  sites[i, 1] <- trials
  sites[i, 2] <- mu1
  sites[i, 3] <- sigma1
  sites[i, 4] <- 1 #site
}

for (i in ((n/3) + 1):((n/3) + (n/3))) {
  mu2 <- rnorm(1, 91, 25) #mu for site 2 stays the same. Vary SD
  sigma2 <- runif(1, 0, 10) #sigma for site 2
  trials <- rnorm(1, mu2, sigma2) #distribution for trials

  sites[i, 1] <- trials
  sites[i, 2] <- mu2
  sites[i, 3] <- sigma2
  sites[i, 4] <- 2 #site
}

for (i in ((n/3) + (n/3) + 1):n) {
  mu3 <- rnorm(1, 91, 45) #mu for site 3 stays the same. Vary SD
  sigma3 <- runif(1, 0, 10) #sigma for site 3
  trials <- rnorm(1, mu3, sigma3) #distribution for trials

  sites[i, 1] <- trials
  sites[i, 2] <- mu3

```

```

    sites[i, 3] <- sigma3
    sites[i, 4] <- 3 #site
  }

  ### THE MODEL - added SITE to sigma!
  dat <- list(trials = as.numeric(sites[, 1]), site = as.integer(sites[,
    4])) #This sets up the data sheet with the simulated data that the model will run on

  s1 <- ulam(alist(trials ~ dnorm(mu[site], sigma[site]), mu[site] ~
    dnorm(91, 21), sigma[site] ~ dunif(0, 40)), data = dat, log_lik = TRUE,
    messages = FALSE) #This is the model

  # Plot it
  labels <- paste("mu[", 1:3, "]:", levels(dat$site), sep = "")
  plot(precis(s1, depth = 2, pars = "mu"), labels = labels, xlab = "expected trials to reverse") #contra

  # Summary data
  precis(s1, depth = 2) #depth=2 means to show any vector parameters (p.156)

  # Compute pairwise contrasts Comparing sites 1 and 2
  post1 <- extract.samples(s1)
  post1$diff_12 <- post1$sigma[, 1] - post1$sigma[, 2] #subtract the SD (in the mu term) for site 2 from
  precis(post1, depth = 2) #look at the last row: diff_12: this is the expected difference between the d
  # hist(post1$diff_12) #another way of visualizing it

  # Comparing sites 1 and 3
  post2 <- extract.samples(s1)
  post2$diff_13 <- post2$sigma[, 1] - post2$sigma[, 3]
  precis(post2, depth = 2)

  # Comparing sites 2 and 3
  post3 <- extract.samples(s1)
  post3$diff_23 <- post3$sigma[, 2] - post3$sigma[, 3]
  precis(post3, depth = 2)

```

394 We then ran the **mathematical model** and performed pairwise contrasts and determined that we will be  
 395 able to detect differences between sites with a sample size of 15 at each site if the average number of trials  
 396 to reverse a preference differs by >13 trials, and the standard deviation is a maximum of 21 at each site  
 397 (Table 2). For a sample size of 20 at each site, which is more like what we expect, we will be able to detect  
 398 site differences if the average number of trials to reverse a preference differs by >11 trials, and the standard  
 399 deviation is a maximum of 23 at each site (Table 2).

400 **Table 2.** Simulation outputs from **varying sample size (n), and  $\alpha$  means and standard deviations.**  
 401 We calculate pairwise contrasts between the estimated means from the posterior distribution: if for a large  
 402 sample the difference is both positive and negative and crosses zero (yes), then we are not able to detect  
 403 differences between the two sites. If the differences between the means are all on one side of zero for 89% of  
 404 the posterior samples (no), then we are able to detect differences between the two sites. We chose the 89%  
 405 interval based on (McElreath 2016). Note that for latency, there is no mu\_sd, but rather one phi that is the  
 406 same for all sites.

```

d <- read.csv(url("https://raw.githubusercontent.com/corinalogan/grackles/master/Files/Preregistrations,
  header = F, sep = ",", stringsAsFactors = F)

colnames(d) <- c("Response variable", "n", "mu1", "mu2", "mu3",
  "mu1_sd", "mu2_sd", "mu3_sd", "Difference crosses zero? S1-S2",

```

```
"Difference crosses zero? S1-S3", "Difference crosses zero? S2-S3",  
"Notes")  
  
library(kableExtra)  
knitr::kable(d) %>% kable_styling(full_width = T, position = "left")
```

Response vari- able	n	mu1	mu2	mu3	mu1_sd	mu2_sd	mu3_sd	Difference crosses zero? S1-S2	Difference crosses zero? S1-S3	Difference crosses zero? S2-S3	Notes	
trials to re- verse	60	91.00	101.00	81.00	21.00	21.00	21.00	Yes	No	No		
trials to re- verse	60	91.00	115.00	81.00	21.00	21.00	21.00	Yes	No	No		
trials to re- verse	60	91.00	120.00	81.00	21.00	21.00	21.00	No	Yes	No		
trials to re- verse	60	91.00	120.00	79.00	21.00	21.00	21.00	No	No	No		
trials to re- verse	60	91.00	120.00	79.00	22.00	22.00	22.00	No	No	No		
trials to re- verse	60	91.00	120.00	79.00	23.00	23.00	23.00	No	No	No		
trials to re- verse	60	91.00	120.00	79.00	24.00	24.00	24.00	No	Yes	No		
trials to re- verse	45	91.00	120.00	79.00	21.00	21.00	21.00	No	Yes	No		
trials to re- verse	45	91.00	120.00	77.00	21.00	21.00	21.00	No	No	No		
trials to re- verse	45	91.00	120.00	77.00	24.00	22.00	23.00	No	Yes	No		
trials to re- verse	45	91.00	120.00	77.00	23.00	22.00	23.00	No	Yes	No		
trials to re- verse	45	91.00	120.00	77.00	22.00	22.00	23.00	Yes	No	No		
trials to re- verse	45	91.00	120.00	77.00	21.00	22.00	23.00	No	Yes	No		
trials to re- verse	45	91.00	120.00	77.00	21.00	21.00	23.00	No	Yes	No		
trials to re- verse	45	91.00	120.00	77.00	22.00	22.00	22.00	No	Yes	No		
trials to re- verse	45	91.00	120.00	77.00	22.00	22.00	21.00	No	Yes	No		
trials to re- verse	45	91.00	120.00	77.00	21.00	22.00	21.00	Yes	No	No		
trials to re- verse	18	91.00	101.00	81.00	21.00	21.00	21.00	Yes	Yes	Yes		
trials to re- verse	15	91.00	101.00	81.00	21.00	21.00	21.00	Yes	Yes	Yes		
trials to re- verse	15	111.00	81.00	71.00	21.00	20	21.00	21.00	Yes	No	Yes	Needs a 40 trial solv-



To investigate the degree to which we can detect differences in the variances between sites, we ran another version of the **mathematical model** using a sample size of 15 per site and we held the mean number of trials to reverse a preference constant between all populations. We reduced the  $\sigma$  standard deviation to 10 (from 40) to reduce the expectation of measurement error, which would overshadow any actual site differences attributable to variation in the standard deviation. We then changed the  $\alpha$  standard deviations and performed pairwise site contrasts. We determined that it will be difficult to detect meaningful differences in variances in the number of trials to reverse a preference between sites (Table 3).

**Table 3.** Simulation outputs from **varying the  $\alpha$  standard deviations (the variance)**, while holding sample size ( $n$ ) and site means ( $\mu$ ) constant. We calculate pairwise contrasts between the estimated standard deviations from the posterior distribution: if for a large sample the difference is both positive and negative and crosses zero (yes), then we are not able to detect differences between the two sites. If the differences between the means are all on one side of zero for 89% of the posterior samples (no), then we are able to detect differences between the two sites. We chose the 89% interval based on (McElreath 2016). Note that for latency, there is no  $\mu\_sd$ , but rather one  $\phi$  that is the same for all sites.

```
d <- read.csv(url("https://raw.githubusercontent.com/corinalogan/grackles/master/Files/Preregistrations",
  header = T, sep = ",", stringsAsFactors = F)

colnames(d) <- c("Response variable", "n", "mu1", "mu2", "mu3",
  "mu1_sd", "mu2_sd", "mu3_sd", "Difference crosses zero? S1-S2",
  "Difference crosses zero? S1-S3", "Difference crosses zero? S2-S3")

library(kableExtra)
knitr::kable(d) %>% kable_styling(full_width = T, position = "left")
```

Response vari- able	n	mu1	mu2	mu3	mu1_sd	mu2_sd	mu3_sd	Difference crosses zero? S1-S2	Difference crosses zero? S1-S3	Difference crosses zero? S2-S3
trials to reverse	45	91	91	91	22	22	22	Yes	Yes	No
trials to reverse	45	91	91	91	15	25	45	No	No	No
trials to reverse	45	91	91	91	1	25	75	No	No	Yes
trials to reverse	45	91	91	91	20	25	32	Yes	No	Yes

### Results (using our actual data)

We will analyze our data using the above model once all of the data have been collected.

### Innovation analysis

#### Model and simulation

Expected values for the number of options solved on the multiaccess log were set to 0-4 (out of 4 options maximum) because this apparatus had been used on two species of jays who exhibited individual variation in the number of loci solved between 0-4 (California scrub-jays and Mexican jays: McCune (2018), Kelsey B McCune et al. (2019)).

$\text{locisolved} \sim \text{Binomial}(4, p)$  [likelihood]

$\text{logit}(p) \sim \alpha[\text{site}]$  [model]

433 locisolved is the number of loci solved on the multiaccess box, 4 is the total number of loci on the multiaccess  
 434 box, p is the probability of solving any one locus across the whole experiment,  $\alpha$  is the intercept, and each  
 435 site gets its own intercept. After running simulations, we identified the following distribution to be the most  
 436 likely priors for our expected data:

437  $\alpha \sim \text{Normal}(0,1)$  [ $\alpha$  prior]

438 We used a normal distribution for  $\alpha$  because it is a sum (see Figure 10.6 in McElreath (2016)) and a logit link  
 439 to ensure the values are between 0 and 1. We set the mean to 0 on a logit scale, which means an individual  
 440 solves 2 loci on average on the actual scale at a probability of 0.5.

```
library(rethinking)
library(rstan)

### SIMULATION
#Figuring out probability, p
s1mean <- 2 #site 1: solves on average 2 loci
s2mean <- 1.6 #site 2: solves on average 1.6 loci
s3mean <- 3.6 #site 3: solves on average 1.6 loci

s1mean <- log((s1mean/4)/(1-(s1mean/4))) #convert to the logit scale: log(p/1-p) p=4/probability to sol
s2mean <- log((s2mean/4)/(1-(s2mean/4)))
s3mean <- log((s3mean/4)/(1-(s3mean/4)))

#Figuring out alpha prior
probabilities<-inv_logit(rnorm(10000,mean=0,sd=0.3))
loci <- vector()
for(i in 1:10000) {loci[i]<-rbinom(1,size=4,prob=probabilities[i])}
mean(loci) #this is the mean number of loci solved
hist(loci) #We expect a bell curve with most of the individuals being average at solving loci (at the m

#Figuring out how much noise we expect (sd in alpha) = 0.4
hist(inv_logit(rnorm(10000,mean=0,sd=1))) #it is spread out along the x axis so the many individuals in
hist(inv_logit(rnorm(10000,mean=0,sd=0.5))) #this is much narrower, thus most of the individuals will m

#Result = in the simulation alpha sd should not be larger than 1 because at 1 it ranges from a probabil

#RUN SIMULATION WITH PARAMETERS (this is the simulated data that the model (below) will use) Generate v

asd <- 0.3 #alpha sd for the population. We set this smaller than what we had it above bc we wanted to
s1mean <- 1.9 #site 1: alpha mean: solves on average 2 loci
s2mean <- 2.1 #site 2: alpha mean: solves on average 1.6 loci
s3mean <- 2.2 #site 3: alpha mean: solves on average 3.6 loci
n <- 60

s1mean <- log((s1mean/4)/(1-(s1mean/4))) #convert to the logit scale: log(p/1-p) p=4/probability to sol
s2mean <- log((s2mean/4)/(1-(s2mean/4)))
s3mean <- log((s3mean/4)/(1-(s3mean/4)))

sites<-matrix(nrow=n,ncol=6)
colnames(sites)<-c("loci","site","probability","logit base value","base probability","logit probability")
for (i in 1:n/3) {
  s1 <- rnorm(1,s1mean,0.5) #n, mean, sd of probability of solving a locus. SD is the noise
  p <- s1
```

```

sites[i,1]<-sum(rbinom(50,1,inv_logit(p)/12.5)) #50 trials (drawing 50 times (the highest number of t
ifelse(sites[i,1]>4,sites[i,1]<-4,sites[i,1]<-sites[i,1])
sites[i,2]<-1 #column: site 1
sites[i,3]<-inv_logit(p) #column: p (probability)
sites[i,4]<-s1 #logit base value of alpha (for the intercept for this site)
sites[i,5]<-inv_logit(s1) #base probability is the inverse logit of the previous column
sites[i,6]<-p #logit probability
}

for (i in ((n/3)+1):((n/3)+(n/3))){
  s2 <- rnorm(1,s2mean,0.5) #n, mean, sd of probability of solving a locus. SD is the noise
  p <- s2
  sites[i,1]<-sum(rbinom(50,1,inv_logit(p)/12.5))
  ifelse(sites[i,1]>4,sites[i,1]<-4,sites[i,1]<-sites[i,1])
  sites[i,2]<-2 #column: site 2
  sites[i,3]<-inv_logit(p)
  sites[i,4]<-s2
  sites[i,5]<-inv_logit(s2)
  sites[i,6]<-p
}

for (i in ((n/3)+(n/3)+1):n){
  s3 <- rnorm(1,s3mean,0.5) #n, mean, sd of probability of solving a locus. SD is the noise
  p <- s3
  sites[i,1]<-sum(rbinom(50,1,inv_logit(p)/12.5))
  ifelse(sites[i,1]>4,sites[i,1]<-4,sites[i,1]<-sites[i,1])
  sites[i,2]<-3 #column: site 3
  sites[i,3]<-inv_logit(p)
  sites[i,4]<-s3
  sites[i,5]<-inv_logit(s3)
  sites[i,6]<-p
}

#Have a look at the relationship to double check the relationship is as expected when back translating
plot(sites[,3],sites[,1]) #positive relationship as expected

### RUN MODEL
#don't change alpha or beta parameters here because they are changed above
dat <- list(locisolved = sites[,1],
            site = sites[,2]
            )

m1 <- ulam( alist(
  locisolved ~ dbinom(4,p) , #4 loci, p=probability of solving a locus
  logit(p) <- a[site] , #site=random effect
  a[site] ~ dnorm(0,1) #each site gets its own intercept, set to sd to 1 to make a slightly narrower di
) , data=dat , chains=4 , log_lik=TRUE )

precis(m1,depth=2)
#mean(sites[1:8,1])
#mean(sites[9:16,1])

```

```

#Figure out alpha sd: check posterior for p to look at the distribution of probabilities that are proba
#prior <- extract.prior(m1,n=1e4)
#p <- inv_logit(prior$a) #convert from logit to actual probability
#dens(p,adj=0.1)
#result1: it was a normal curve with a peak at 0.5

#Plot it
labels <- paste( "a[" , 1:3 , "]:" , levels(dat$site) , sep="" )
plot( precis( m1 , depth=2 , pars="a" ) , labels=labels , xlab="expected loci solved" ) #contrasts plo

#Compute pairwise contrasts
#Comparing sites 1 and 2
post1 <- extract.samples(m1)
post1$diff_12 <- post1$a[,1] - post1$a[,2] #subtract mu for site 2 from mu for site 1. col1=site 1, col
precis(post1,depth=2) #look at the last row: diff_12: this is the expected difference between the distr

#Comparing sites 1 and 3
post2 <- extract.samples(m1)
post2$diff_13 <- post2$a[,1] - post2$a[,3]
precis(post2,depth=2)

#Comparing sites 2 and 3
post3 <- extract.samples(m1)
post3$diff_23 <- post3$a[,2] - post3$a[,3]
precis(post3,depth=2)

```

We then ran the **mathematical model** and performed pairwise contrasts and determined that we will be able to detect differences between sites with a sample size of 15 at each site if the average number of loci solved differs by 1.2 loci or more and the standard deviation is generally a maximum of 0.9 at each site (Table 2). For a sample size of 20 at each site, we will be able to detect site differences if the average number of loci solved differs by 0.7 of a locus or more and the standard deviation is generally a maximum of 1 at each site (Table 2). Note: the Arizona sample size is 11 for the multiaccess log and 17 on a similar multiaccess box.

Because the mean and the variance are linked in the binomial distribution, and because the variance simulations in the flexibility analysis showed that we will not be able to robustly detect differences in variance between sites, we will plot the variance in the number of loci solved between sites to determine whether the edge population has a wider or narrower spread than the other two populations.

### Results (using our actual data)

We will analyze our data using the above model once all of the data have been collected.

### Exploration analysis

#### Model and simulation

We modeled the average latency to approach an object and compared these between sites. We simulated data and set the model as follows:

latency  $\sim$  gamma-Poisson( $\lambda_i, \phi$ ) [likelihood]

$\log(\lambda_i) \sim \alpha[\text{site}]$  [the model]

latency is the average latency to approach an object,  $\lambda_i$  is the rate (probability of approaching the object in each second) per bird (and we take the log of it to make sure it is always positive; birds with a higher rate have a smaller latency),  $\phi$  is the dispersion of the rates across birds, and  $\alpha$  is the intercept for the rate per site.

Expected values for the latency to approach a novel object range from 0-2700 sec, which encompasses the time period during which they are exposed to the object (sessions last up to 45 min). However, we do not provide an upper limit for the model because those birds that do not approach within 2700 sec would eventually have had to approach the object to access their food (it is just that sessions did not run that long). After running simulations, we identified the following distribution and priors to be the most likely for our expected data:

$\phi \sim 1/(\text{Exponential}(1))$  [ $\phi$  prior]

$\alpha \sim \text{Normal}(1350, 500)$  [ $\alpha$  prior]

We used a gamma-Poisson distribution for latency because it constrains the values to be positive. For  $\phi$ , we used an exponential distribution because it is standard for this parameter. We used a normal distribution for  $\alpha$  because it is a sum with a large mean (see Figure 10.6 in McElreath (2016)). We estimate that the grackles might approach the object at any time in the session, therefore we held the  $\alpha$  mean of 1350 sec in mind as we conducted the modeling. We set the  $\alpha$  standard deviation to 500 because this puts the range of seconds for the distribution in the possible range.

```
library(rethinking)
library(rstan)
library(Rcpp)

# Get citations for packages
options(citation.bibtex.max = 999) #this makes it so that citation() only shows up in bibtex format
citation("Rcpp")
citation("rstan")
citation("rethinking")

# Checking to make sure the alpha prior is as expected
hist(rnorm(10000, mean = 1350, sd = 500)) #this gives us a range from 0-3000 sec, which is what is pos.

# Looked into using a poisson rather than a gamma poisson
# distribution, but poisson only has one parameter (lambda)
# and it doesn't let me change the shape of the curve. So I
# need to use a gamma poisson distribution
hist(rpois(1000, 1350)) #1000=sample size, mean=1350 sec

### SIMULATIONS: GAMMA-POISSON NOTE: we varied the sample size
### and mean to explore the boundaries of differences we would
### be able to detect between sites. However, we did not vary
### the variance because in this equation, the mean and
### variance are tied together. FYI this is a useful resource
### for gamma poisson distributions
### https://en.wikipedia.org/wiki/Gamma_distribution Lambda *
### phi = variance. NOTE: lambda=mu, and phi is sometimes
### called scaling factor or scale

# Find the right mean seconds to approach (alpha mean) and
# phi
hist(rgammpois(10000, 2000, 1000), breaks = 30) #10000 samples, mu (average seconds to approach), scaling factor
sum(rgammpois(10000, 2000, 1000) > 2700) #use this to see how many of the 1000 simulated grackles did not approach
median(rgammpois(10000, 2000, 1000)) #check the median to see if it is around 1350 sec. =1652
var(rgammpois(10000, 2000, 1000)) #variance of the 1000 individuals in the sample above, mean=2000 seconds
# Result: log(lambda)=alpha = 2000 seconds (this is alpha
# mu), phi=1000
```

```

# Simulate the data
n <- 60 #number of individuals
phi <- 1000 #this gives us a distribution that we expect (long right tail)

am1 <- 4.6 #site 1 alpha mean: 4.6=log of 100 sec (seconds to approach an object), 110s=4.7
am2 <- 6.3 #site 2 alpha mean: 6.2=log of 500 sec, 550s=6.3, 600s=6.4
am3 <- 7.3 #site 3 alpha mean: 7.1=log 1200 sec, 1339s=7.2, 1480s=7.3, 1636s=7.4

sites <- matrix(nrow = n, ncol = 2)
colnames(sites) <- c("latency", "site")

for (i in 1:n/3) {
  lambda <- exp(am1) #the linear model for site 1
  latency <- rgam pois(1, lambda, phi) #this is the latency per individual
  sites[i, 1] <- latency
  sites[i, 2] <- 1 #site 1
}

for (i in ((n/3) + 1):((n/3) + (n/3))) {
  lambda <- exp(am2)
  latency <- rgam pois(1, lambda, phi)
  sites[i, 1] <- latency
  sites[i, 2] <- 2 #site 2
}

for (i in ((n/3) + (n/3) + 1):n) {
  lambda <- exp(am3)
  latency <- rgam pois(1, lambda, phi)
  sites[i, 1] <- latency
  sites[i, 2] <- 3 #site 3
}

sites[sites[, 1] > 2700, 1] <- 2701 #make all non-participatory birds have a 2701

# In case I want to check what the data look like
# mean(sites[sites[,2]==1,1]) #site 1
# mean(sites[sites[,2]==2,1]) #site 2
# mean(sites[sites[,2]==3,1]) #site 3

# hist(sites[sites[,2]==1,1],breaks=30) #site 1
# hist(sites[sites[,2]==2,1],breaks=30) #site 2
# hist(sites[sites[,2]==3,1],breaks=30) #site 1

# Maybe we want to compare the proportion of observations
# below the median between 2 sites?
# median(sites[sites[,2]!=3,1]) # 62
# sum(sites[sites[,2]==1,1]>62) # 7
# sum(sites[sites[,2]==2,1]>62) # 13

### RUN THE MODEL
dat <- list(latency = as.integer(sites[, 1]), site = as.integer(sites[,
2]))

```



```

m2 <- ulam(alist(latency ~ dgamma(lambda, phi), log(lambda) <- a[site],
  a[site] ~ dnorm(1, 1), phi ~ dexp(1)), data = dat, log_lik = TRUE,
  messages = FALSE)

### NOTE: this model doesn't work on my computer...something
### about the rcpp package needing to be installed from
### scratch? Follow instructions here:
### https://github.com/stan-dev/rstan/wiki/How-to-build-rstan-package%3F

# Plot it (contrasts plot p.157)
labels <- paste("a[", 1:3, "]:", levels(dat$site), sep = "")
plot(precis(m2, depth = 2, pars = "a"), labels = labels, xlab = "expected latency")

# Compute pairwise contrasts Comparing sites 1 and 2
post1 <- extract.samples(m2)
post1$diff_12 <- post1$a[, 1] - post1$a[, 2] #subtract mu for site 2 from mu for site 1. col1=site 1,
precis(post1, depth = 2) #look at the last row: diff_12: this is the expected difference between the d

# Comparing sites 1 and 3
post2 <- extract.samples(m2)
post2$diff_13 <- post2$a[, 1] - post2$a[, 3]
precis(post2, depth = 2)

# Comparing sites 2 and 3
post3 <- extract.samples(m2)
post3$diff_23 <- post3$a[, 2] - post3$a[, 3]
precis(post3, depth = 2)

```

We then ran the **mathematical model** and performed pairwise contrasts and determined that we will be able to detect differences between sites with a sample size of 15 at each site or 20 at each site if the average latency to approach the object differs by at least 450 sec at each site (Table 2). We kept the shape of the curve (which can be thought of as similar to a standard deviation or the variance) the same across sites because we do not think this assumption will change across populations (i.e., there will be lots of variation at each site with some individuals approaching almost immediately, others in the middle of the session, and others near the end).

Because the mean and the variance are linked in the gamma-Poisson distribution, and because the variance simulations in the flexibility analysis showed that we will not be able to robustly detect differences in variance between sites, we will plot the variance in the latency to approach the object between sites to determine whether the edge population has a wider or narrower spread than the other two populations.

### Results (using our actual data)

We will analyze our data using the above model once all of the data have been collected.

## Persistence analysis

### Model and simulation

Expected values for the number of trials not participated in could range from 0-125 (likely maxima: 300 trials reversal learning [70 trials initial discrimination, 130 trials reversal, ~100 non-participation trials], 50 trials multiaccess log [~25 non-participation trials]). After running simulations, we identified the following distribution and priors most likely for our expected data:

participated ~ Binomial(totaltrials, p) *[likelihood]*

497  $\text{logit}(p) \sim \alpha[\text{site}]$  *[model]*

498 participated indicates whether the bird participated or not in a given trial, total trials is the total number  
 499 of trials offered to the individual (those participated in plus those not participated in), p is the probability  
 500 of participating in a trial,  $\alpha$  is the intercept, and each site gets its own intercept. We used a logit link to  
 501 constrain the output to between 0 and 1. After running simulations, we identified the following distribution  
 502 and priors most likely for our expected data:

503  $\alpha \sim \text{Normal}(0, 0.5)$  *[\alpha prior]*

504 We used a normal distribution for  $\alpha$  because it is a sum (see Figure 10.6 in McElreath (2016)). We set the  
 505 mean to 0 (on a logit scale, which is a probability of 0.5 that a bird will participate in every other trial on  
 506 average on the actual scale).

```
library(rethinking)
library(rstan)
library(Rcpp)

# SIMULATE DATA Make sure the simulation stays in the correct
# range
hist(rnorm(n, mean = 275, sd = 27)) #mean number of total trials, based on the estimation that they wi

# Figure out the priors: alpha mean
hist(inv_logit(rnorm(2000, mean = 0, sd = 1.5))) #what do we expect the probability of participating t

# Manipulate these variables
n <- 45 #number of grackles in total across all 3 sites
site1mean <- 0.7 #site 1 alpha mean: proportion of trials participated in
site2mean <- 0.7
site3mean <- 0.7
asd1 <- 0.25 #site 1 alpha sd: proportion of trials participated in
asd2 <- 0.25 #site 2 alpha sd
asd3 <- 0.25 #site 3 alpha sd

# Populate the simulated data sheet
testdata <- matrix(nrow = n, ncol = 5)
testdata <- as.data.frame(testdata)
colnames(testdata) <- c("individual", "site", "participated",
  "optout", "totaltrials")
testdata$individual <- c(1:n) #unique identifier for each individual
testdata$site <- c(rep(1, n/3), rep(2, n/3), rep(3, n/3)) #assign 3 sites
testdata$totaltrials <- c(round(rnorm(n, mean = 275, sd = 27),
  0)) #mean number of total trials=275 at each site with some variation (sd=15) among individuals, b

## Simulate the number of trials participated in for each
## site. Run a loop to multiply the proportion participated in
## by the total trials to convert it to the number of trials
## participated in per individual
for (i in 1:(n/3)) {
  probability <- rnorm(1, site1mean, asd1)
  ifelse(probability > 1, probability <- 1, probability <- probability)
  testdata[i, ]$participated <- round(testdata[i, ]$totaltrials *
    probability, 0)
}

for (i in ((n/3) + 1):(2 * n)/3) {
```

```

probability <- rnorm(1, site2mean, asd2)
ifelse(probability > 1, probability <- 1, probability <- probability)
testdata[i, ]$participated <- round(testdata[i, ]$totaltrials *
  probability, 0)
}

for (i in ((2 * n/3) + 1):n) {
  probability <- rnorm(1, site3mean, asd3)
  ifelse(probability > 1, probability <- 1, probability <- probability)
  testdata[i, ]$participated <- round(testdata[i, ]$totaltrials *
    probability, 0)
}

testdata$optout <- testdata$totaltrials - testdata$participated #Populate the optout column: Calculate

# RUN THE MODEL (this is the simulated data that the model
# (below) will use) Generate values for simulated individuals
# from 3 simulated sites: vary the alpha site mean (expected
# number of trials participated in and whether sites differ)
# and alpha sd (the noise/variance in the population in the
# number of trials participated in), and sample size
dat_list <- list(participated = as.integer(testdata$participated),
  totaltrials = as.integer(testdata$totaltrials), site = as.integer(testdata$site))

m1 <- ulam(alist(participated ~ dbinom(totaltrials, p), logit(p) <- a[site],
  a[site] ~ dnorm(0, 1.5)), data = dat_list, chains = 4)

# Plot it
post <- extract.samples(m1)
probability <- inv_logit(post$a)
labels <- paste("a[", 1:3, "]:", levels(dat_list$site), sep = "")
plot(precis(as.data.frame(probability)), xlim = c(0, 1), labels = labels,
  xlab = "proportion of trials participated in") #contrasts plot p.157

# Summary data
precis(m1, depth = 2) #depth=2 means to show any vector parameters (p.156)

# Compute pairwise contrasts Comparing sites 1 and 2
post1 <- extract.samples(m1)
post1$diff_12 <- inv_logit(post1$a[, 1]) - inv_logit(post1$a[,
  2]) #subtract mu for site 2 from mu for site 1. col1=site 1, col2=site 2
precis(post1, depth = 2) #look at the last row: diff_12: this is the expected difference between the d

# Comparing sites 1 and 3
post2 <- extract.samples(m1)
post2$diff_13 <- inv_logit(post2$a[, 1]) - inv_logit(post2$a[,
  3])
precis(post2, depth = 2)

# Comparing sites 2 and 3
post3 <- extract.samples(m1)
post3$diff_23 <- inv_logit(post3$a[, 2]) - inv_logit(post3$a[,
  3])

```

```
precis(post3, depth = 2)
```

We then ran the **mathematical model** and performed pairwise contrasts and determined that we will be able to detect differences between sites with a sample size of 15 per site or 20 per site if the average proportion of trials participated in differs by at least 0.08 and the standard deviation is generally a maximum of 0.25 at each site (Table 2).

Because the mean and the variance are linked in the binomial distribution, and because the variance simulations in the flexibility analysis showed that we will not be able to robustly detect differences in variance between sites, we will plot the variance in the proportion of trials participated in between sites to determine whether the edge population has a wider or narrower spread than the other two populations.

### Results (using our actual data)

We will analyze our data using the above model once all of the data have been collected.

### Repeatability of exploration and persistence

**Analysis:** We will obtain repeatability estimates that account for the observed and latent scales, and then compare them with the raw repeatability estimate from the null model. The repeatability estimate indicates how much of the total variance, after accounting for fixed and random effects, is explained by individual differences (bird ID). We will run this GLMM using the MCMCglmm function in the MCMCglmm package ((Hadfield 2010)) with a Poisson distribution and log link using 13,000 iterations with a thinning interval of 10, a burnin of 3,000, and minimal priors ( $V=1$ ,  $\nu=0$ ) (Hadfield 2014). We will ensure the GLMM shows acceptable convergence (i.e., lag time autocorrelation values  $<0.01$ ; (Hadfield 2010)), and adjust parameters if necessary.

```
# Persistence
per <- read.csv(url(""), header = T, sep = ",", stringsAsFactors = F)

# DATA CHECKING
library(DHARMA)
library(lme4)
simulationOutput <- simulateResiduals(fittedModel = glmer(NoParticipation ~
  Test + (1 | ID), family = poisson, data = per), n = 250)
simulationOutput$scaledResiduals
testDispersion(simulationOutput)
testZeroInflation(simulationOutput)
testUniformity(simulationOutput)
plot(simulationOutput)
plotResiduals(Test, simulationOutput$scaledResiduals) #can't get this code to work yet

# REPEATABILITY GLMM
library(MCMCglmm)
prior = list(R = list(R1 = list(V = 1, nu = 0)), G = list(G1 = list(V = 1,
  nu = 0)))
pers <- MCMCglmm(NoParticipation ~ Test, random = ~ID, family = "poisson",
  data = per, verbose = F, prior = prior, nitt = 13000, thin = 10,
  burnin = 3000)
summary(pers)
# autocorr(pers$Sol) #Did fixed effects converge?
# autocorr(pers$VCV) #Did random effects converge?

# In MCMCglmm, the latent scale adjusted repeatability and
# its credible interval can simply be obtained by:
# mod$VCV[,ID]/(mod$VCV[,ID]+mod$VCV[,units]) - advice from
```

```
# Maxime Dahirel
```

```
repeata <- pers$VCV[, "ID"]/(pers$VCV[, "ID"] + pers$VCV[, "units"]) #latent scale adjusted repeatability  
mean(repeata)  
var(repeata)  
posterior.mode(repeata)  
HPDinterval(repeata, 0.95)
```

```
# Repeatability on the data/observed scale (accounting for  
# fixed effects) code from Supplementary Material S2 from  
# Villemereuil et al. 2018 J Evol Biol
```

```
vf <- sapply(1:nrow(pers[["Sol"]]), function(i) {  
  var(predict(pers, it = i))  
}) #estimates for each iteration of the MCMC
```

```
repeataF <- (vf + pers$VCV[, "ID"])/(vf + pers$VCV[, "ID"] +  
  pers$VCV[, "units"]) #latent scale adjusted + data scale  
posterior.mode(repeataF)  
HPDinterval(repeataF, 0.95)
```

```
# Now compare with the raw repeatability: null model
```

```
persraw <- MCMCglmm(NoParticipation ~ 1, random = ~ID, family = "poisson",  
  data = per, verbose = F, prior = prior, nitt = 13000, thin = 10,  
  burnin = 3000)  
summary(persraw)
```

```
repeataraw <- persraw$VCV[, "ID"]/(persraw$VCV[, "ID"] + persraw$VCV[,  
  "units"]) #latent scale adjusted repeatability and its credible interval  
posterior.mode(repeata)  
HPDinterval(repeata, 0.95)
```

```
# Exploration of novel object
```

```
eo <- read.csv(url(""), header = T, sep = ",", stringsAsFactors = F)
```

```
# DATA CHECKING
```

```
library(DHARMA)
```

```
library(lme4)
```

```
simulationOutput <- simulateResiduals(fittedModel = glmer(Latency ~  
  Time + (1 | ID), family = poisson, data = eo), n = 250) #250 simulations, but if want higher precision  
simulationOutput$scaledResiduals #Expect a flat distribution of the overall residuals, and uniformity  
testDispersion(simulationOutput) #if under- or over-dispersed, then p-value<0.05, but then check the data  
testZeroInflation(simulationOutput) #compare expected vs observed zeros, not zero-inflated if p<0.05  
testUniformity(simulationOutput) #check for heteroscedasticity ('a systematic dependency of the dispersion)  
plot(simulationOutput) #...there should be no pattern in the data points in the right panel  
plotResiduals(Test, simulationOutput$scaledResiduals) #plot the residuals against other predictors (in
```

```
# REPEATABILITY GLMM
```

```
library(MCMCglmm)
```

```
prior = list(R = list(R1 = list(V = 1, nu = 0)), G = list(G1 = list(V = 1,  
  nu = 0)))
```

```
explo <- MCMCglmm(Latency ~ Time, random = ~ID, family = "poisson",  
  data = eo, verbose = F, prior = prior, nitt = 13000, thin = 10,  
  burnin = 3000)  
summary(pers)
```

```

# autocorr(explo$Sol) #Did fixed effects converge?
# autocorr(explo$VCV) #Did random effects converge?

# In MCMCglmm, the latent scale adjusted repeatability and
# its credible interval can simply be obtained by:
# mod$VCV[,ID]/(mod$VCV[,ID]+mod$VCV[,units]) - advice from
# Maxime Dahiriel

repeata <- explo$VCV[, "ID"]/(explo$VCV[, "ID"] + explo$VCV[,
  "units"]) #latent scale adjusted repeatability and its credible interval
mean(repeata)
var(repeata)
posterior.mode(repeata)
HPDinterval(repeata, 0.95)

# Repeatability on the data/observed scale (accounting for
# fixed effects) code from Supplementary Material S2 from
# Villemereuil et al. 2018 J Evol Biol
vf <- sapply(1:nrow(explo[["Sol"]]), function(i) {
  var(predict(explo, it = i))
}) #estimates for each iteration of the MCMC

repeataF <- (vf + explo$VCV[, "ID"])/(vf + explo$VCV[, "ID"] +
  explo$VCV[, "units"]) #latent scale adjusted + data scale
posterior.mode(repeataF)
HPDinterval(repeataF, 0.95)

# Now compare with the raw repeatability: null model
exploraw <- MCMCglmm(Latency ~ 1, random = ~ID, family = "poisson",
  data = eo, verbose = F, prior = prior, nitt = 13000, thin = 10,
  burnin = 3000)
summary(exploraw)

repeataraw <- exploraw$VCV[, "ID"]/(exploraw$VCV[, "ID"] + exploraw$VCV[,
  "units"]) #latent scale adjusted repeatability and its credible interval
posterior.mode(repeata)
HPDinterval(repeata, 0.95)

```

## Q2: dispersal

### Response variable

- 1) Average relatedness between all pairs of individuals within one sex

### Explanatory variables

- 1) Site diameter (meters)
- 2) Site sample size
- 3) Number of generations at a site

One model will be run per sex

The data will be analyzed as in Sevchik et al. (2019). To summarize, blood is collected from the bird, DNA is extracted (by Aaron Blackwell at Washington State University), size selected (between 400-700 base pairs), and sequenced using ddRADseq (at Cornell University Lab of Ornithology) on an Illumina NextSeq500



machine using the mid-output setting for 150 base pair single end reads. Data are post processed to generate single nucleotide polymorphisms (SNPs) as in Thrasher et al. (2018). Genetic relatedness between all pairs of individuals is calculated using the package “related” (Pew et al. (2015)) in R (as in Thrasher et al. (2018)) using the estimator by Queller & Goodnight, which was more robust for our inferences in a subset of the Arizona data (Sevchik et al. 2019). Permutations (i.e., randomly assigning site ID to individuals) and general linear models estimating average relatedness of each individual to all others at that site will be used to determine whether individuals at one site are more closely related to each other than the individuals at another site.

## Model and simulation

Expected values for average relatedness per bird were based on the fact that average relatedness with these estimators has to range between -1 and 1 and because it is an average we expect a normal distribution.

$\text{averagerelatedness} \sim \alpha[\text{site}]$  [the model]

$\alpha[\text{site}] \sim \text{Normal}(\mu, \sigma)$  [ $\alpha_1$  prior]

$\mu \sim \text{Normal}(0, 1)$  [ $\mu$  prior]

$\sigma \sim \text{Uniform}(0, 1)$  [ $\sigma$  prior]

## Q3: habitat

**P3: GTGR & BTGR use different habitats and GTGR’s habitat has increased over time and P4: GTGR increased habitat breadth over time, but BTGR did not**

*Response variable:* Presence/absence of GTGR and BTGR

*Explanatory variable*

- 1) **Land cover** (e.g., forest, urban, arable land, pastureland, wetlands, marine coastal, grassland, man-grove) - we chose these land cover types because they represent the habitat types in which both species exist, as well as habitat types (e.g., forest) they are not expected to exist in (Selander and Giller 1961) to confirm that this is the case. If it is the case, it is possible that large forested areas are barriers for the range expansion of one or both species. We will download global land cover type data from MODIS (16 terrestrial habitat types) and/or the IUCN habitat classification (47 terrestrial habitat types). The IUCN has assigned habitat classifications to great-tailed (<https://www.iucnredlist.org/species/22724308/132174807#habitat-ecology>) and boat-tailed (<https://www.iucnredlist.org/species/22724311/94859792#habitat-ecology>) grackles, however these appear to be out of date and we will update them for the purposes of this project.
- 2) **Elevation** - Selander and Giller (1961) notes the elevation range for GTGR (0-2134m), but not BTGR, therefore establishing the current elevation ranges for both species will allow us to determine whether and which mountain ranges present range expansion challenges. We will obtain elevation data from USGS.
- 3) **Climate** (e.g., daily/annual temperature range) - because this species was originally tropical (Wehtje 2003), which generally has a narrow daily and annual climate range, and now they exist in temperate regions, which have much larger climate ranges, this variable will allow us to determine potential climatic limits for both species. If there are limits, this could inform the difference between the range expansion rates of the two species. We will consider the 19 bioclimatic variables from WorldClim.
- 4) **Presence/absence of water in the cell for each point** - both species are considered to be highly associated with water (e.g., Selander and Giller 1961), therefore we will identify how far from water each species can exist to determine whether it is a limiting factor in the range expansion of one or both species. The data will come from USGS National Hydrography.
- 5) **Connectivity:** Distance between points on the northern edge of the range to the nearest uninhabited suitable habitat patch to the north in 1970 compared with the same patches in ~2018. We identified the northern edge of the distribution based on reports on eBird.org from 1968-1970, which resulted

in recordings of GTGR in 48 patches and recordings of BTGR in 30 patches. For these patches, we calculated the connectivity (the least cost path) to the nearest uninhabited suitable habitat patch in 1970 and again in ~2018. Given that GTGR are not found in forests and that the elevation limits for GTGR (Selander and Giller 1961), and observing the sightings of both species on eBird.org, large forests, tall mountain ranges and high elevation geographic features could block or slow the expansion of one or both species into these areas and their surroundings. For each point, we will calculate the least cost path between it and the nearest location with grackle presence using the leastcostpath R package (Lewis (2020)). This will allow us to determine the costs involved in a grackle deciding whether to fly around or over a mountain range/forest. We will define the forest and mountain ranges from the land cover and/or elevation maps.

One model, including all explanatory variables, will be run for GTGR and a separate model will be run for BTGR. For the explanatory variables, MaxEnt produces a continuous prediction of habitat suitability for each grid cell (0 is least suitable and 1 is most suitable). We will also use jackknifing procedures to evaluate the relative contribution/importance of different environmental variables to the probability of species occurrence. We will optimize the model by trying different regularization coefficient values, which controls how much additional terms are penalized (Maxent's way of protecting against overfitting), and choosing the value that maximizes model fit. Most MaxEnt papers use cross-validation and the area under the curve (AUC) to evaluate model performance, and we will do the same.

### Analysis instructions

- 1) Download and preprocess eBird data. Conduct spatial filtering to account for sampling bias
- 2) Clean the species occurrence data: remove any uncertain records or geographic outliers
- 3) Import climatic variables from WorldClim and landscape data from MODIS and crop to region of interest
- 4) Match environmental data to grackle occurrence records
- 5) Fit models with maxent to get predicted distributions and estimate importance/contribution of each environmental variable

We will refer to Strimas-Mackey et al. (2016) best practices for using eBird data when extracting data on grackle presence in a region from eBird.org. We will gather environmental data from databases, including a database that maps global urban change from 1985-2015 to a high (30 m) resolution (Liu et al. (2020)). We will use a variety of R packages, including auk (Strimas-Mackey et al. (2018)), dismo (Hijmans et al. (2017)), raster (Hijmans (2020)), maptools (Bivand and Lewin-Koh (2019)), tidyverse (Wickham et al. (2019)), rgdal (Bivand et al. (2019)), rJava (Urbanek (2020)), and elevatr (Hollister and Tarak Shah (2017)).

**Analysis 1 (P3: different habitats):** does the range of variables that characterize suitable habitat for GTGR differ from that of BTGR? We will fit species distribution models for both species in 2018 to identify the variables that characterize suitable habitat. We will examine the raw distributions of these variables from known grackle occurrence points or extract information on how the predicted probability of grackle presence changes across the ranges for each habitat variable. The habitat variables for each species will be visualized in a figure that shows the ranges of each variable and how much the ranges of the variables overlap between the two species or not.

**Analysis 2 (P3: habitat suitability):** has the available habitat for both species increased over time? We will fit species distribution models for both species in 1970 and in 2018 and determine for each variable, the range in which grackles are present (we define this as the habitat suitability for each species). Then we will take these variables and identify which locations in the Americas fall within the grackle-suitable ranges in 1970 and in 2018. We will then be able to compare the maps (1970 and 2018) to determine whether the amount of suitable habitat has increased or decreased.

If we are able to find data for these variables before 1970 across the Americas, we will additionally run models using the oldest available data to estimate the range of suitable habitat earlier in their range expansion.

**Analysis 3 (P3: habitat connectivity):** has the habitat connectivity for both species increased over time? If the connectivity distances are smaller in 2018, this will indicate that habitat connectivity has increased over time. We will calculate the least cost path from the northern edge to the nearest suitable

habitat patch. To compare the distances between 1970 and 2018, and between the two species, we will run two models where both have the distance as the response variable and a random effect of location to match the location points over time. The explanatory variable for model 1 will be the year (1970, 2018), and for model 2 it will be the species (GTGR, BTGR).

If we are able to find data for these variables before 1970 across the Americas, we will additionally run models using the oldest available data to estimate the range of connected habitat earlier in their range expansion.

**Analysis 4 (P4: habitat breadth):** has the habitat breadth of both species changed over time? We will count the number of different land cover categories each species is or was present in for 1970 and 2018. To determine whether this influences their distributions, we will calculate how much area in the Americas is in each land cover category, which would then indicate how much habitat is suitable (based solely on land cover) for each species.

```
# Load packages
library(auk)
library(dismo)
library(raster)
library(maptools)
library(tidyverse)
library(rgdal)
library(rJava)

# Get citations for packages
options(citation.bibtex.max = 999) #this makes it so that citation() only shows up in bibtex format
citation("elevatr")
citation("auk")
citation("dismo")
citation("raster")
citation("maptools")
citation("tidyverse")
citation("rgdal")
citation("rJava")

d <- read.csv("gtgr_points.csv", header = T)
```

## F. ETHICS

This research is carried out in accordance with permits from the:

- 1) US Fish and Wildlife Service (scientific collecting permit number MB76700A-0,1,2)
- 2) US Geological Survey Bird Banding Laboratory (federal bird banding permit number 23872)
- 3) Arizona Game and Fish Department (scientific collecting license number SP594338 [2017], SP606267 [2018], SP639866 [2019], and SP402153 [2020])
- 4) Institutional Animal Care and Use Committee at Arizona State University (protocol number 17-1594R)
- 5) California Department of Fish and Wildlife (scientific collecting permit [specific use] number S-192100001-19210-001)

## G. AUTHOR CONTRIBUTIONS

**Logan:** Hypothesis development, data collection (Q1 & Q2), data analysis and interpretation, write up, revising/editing, materials/funding.

**McCune:** Method development, data collection (Q1 & Q2), data analysis and interpretation, revising/editing.

**Breen:** Data collection (Q1 & Q2), data analysis and interpretation, revising/editing.

**Chen (Q3):** Hypothesis development, data collection, ecological niche modeling, data interpretation, revising/editing.

**Lukas:** Hypothesis development, data analysis and interpretation, write up, revising/editing.

## H. FUNDING

This research is funded by the Department of Human Behavior, Ecology and Culture at the Max Planck Institute for Evolutionary Anthropology.

## I. CONFLICT OF INTEREST DISCLOSURE

We, the authors, declare that we have no financial conflicts of interest with the content of this article. CJ Logan and D Lukas are Recommenders at PCI Ecology, and CJ Logan is on the Managing Board at PCI Ecology.

## J. ACKNOWLEDGEMENTS

We thank Kristine Johnson for technical advice on great-tailed grackles; Julia Cissewski and Sophie Kaube for tirelessly solving problems involving financial transactions and contracts; and Richard McElreath for project support.

## K. REFERENCES

- Auersperg AMI, Bayern AMP von, Gajdon GK, Huber L, Kacelnik A. 2011. Flexibility in problem solving and tool use of kea and New Caledonian crows in a multi access box paradigm. *PLOS ONE*. 6(6):e20231. doi:10.1371/journal.pone.0020231. [accessed 2017 May 15]. <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0020231>.
- Auersperg AM, Szabo B, Von Bayern AM, Kacelnik A. 2012. Spontaneous innovation in tool manufacture and use in a goffin's cockatoo. *Current Biology*. 22(21):R903–R904.
- Bird CD, Emery NJ. 2009. Insightful problem solving and creative tool modification by captive nontool-using rooks. *Proceedings of the National Academy of Sciences*. 106(25):10370–10375.
- Bivand R, Keitt T, Rowlingson B. 2019. Rgdal: Bindings for the 'geospatial' data abstraction library. <https://CRAN.R-project.org/package=rgdal>.
- Bivand R, Lewin-Koh N. 2019. Maptools: Tools for handling spatial objects. <https://CRAN.R-project.org/package=maptools>.
- Bolker BM. 2008. *Ecological models and data in r*. Princeton University Press.
- Carpenter B, Gelman A, Hoffman MD, Lee D, Goodrich B, Betancourt M, Brubaker M, Guo J, Li P, Riddell A. 2017. Stan: A probabilistic programming language. *Journal of statistical software*. 76(1).
- Chejanovski ZA, Avilés-Rodríguez KJ, Lapiedra O, Preisser EL, Kolbe JJ. 2017. An experimental evaluation of foraging decisions in urban and natural forest populations of anolis lizards. *Urban Ecosystems*. 20(5):1011–1018.
- Chow PKY, Lea SE, Leaver LA. 2016. How practice makes perfect: The role of persistence, flexibility and learning in problem-solving efficiency. *Animal behaviour*. 112:273–283.
- Ciani AC. 1986. Intertroop agonistic behavior of a feral rhesus macaque troop ranging in town and forest areas in india. *Aggressive behavior*. 12(6):433–439.
- Collias EC, Collias NE. 1964. The development of nest-building behavior in a weaverbird. *The Auk*. 81(1):42–52.
- Eddelbuettel D, François R. 2011. Rcpp: Seamless R and C++ integration. *Journal of Statistical Software*. 40(8):1–18. doi:10.18637/jss.v040.i08. <http://www.jstatsoft.org/v40/i08/>.

700 Federspiel IG, Garland A, Guez D, Bugnyar T, Healy SD, Güntürkün O, Griffin AS. 2017. Adjusting  
701 foraging strategies: A comparison of rural and urban common mynas (*acridotheres tristis*). *Animal cognition*.  
702 20(1):65–74.

703 Gamer M, Lemon J, Gamer MM, Robinson A, Kendall’s W. 2012. Package ‘irr’. Various coefficients of  
704 interrater reliability and agreement.

705 Goldewijk KK. 2001. Estimating global land use change over the past 300 years: The hyde database. *Global*  
706 *biogeochemical cycles*. 15(2):417–433.

707 Griffin AS, Guez D. 2014. Innovation and problem solving: A review of common mechanisms. *Behavioural*  
708 *Processes*. 109:121–134.

709 Hadfield J. 2014. MCMCglmm course notes. [http://cran.r-project.org/web/packages/MCMCglmm/  
710 vignettes/CourseNotes.pdf](http://cran.r-project.org/web/packages/MCMCglmm/vignettes/CourseNotes.pdf).

711 Hadfield JD. 2010. MCMC methods for multi-response generalized linear mixed models: The MCMCglmm  
712 R package. *Journal of Statistical Software*. 33(2):1–22. <http://www.jstatsoft.org/v33/i02/>.

713 Hanski I, Gilpin M. 1991. Metapopulation dynamics: Brief history and conceptual domain. *Biological*  
714 *journal of the Linnean Society*. 42(1-2):3–16.

715 Hijmans RJ. 2020. Raster: Geographic data analysis and modeling. [https://CRAN.R-project.org/package=  
716 raster](https://CRAN.R-project.org/package=raster).

717 Hijmans RJ, Phillips S, Leathwick J, Elith J. 2017. Dismo: Species distribution modeling. [https://CRAN.R-  
718 project.org/package=dismo](https://CRAN.R-project.org/package=dismo).

719 Hollister J, Tarak Shah. 2017. Elevatr: Access elevation data from various apis. [http://github.com/usepa/  
720 elevatr](http://github.com/usepa/elevatr).

721 Hutcheon JA, Chiolero A, Hanley JA. 2010. Random measurement error and regression dilution bias. *Bmj*.  
722 340:c2289.

723 International B. 2018. *Quiscalus mexicanus*. The IUCN Red List of Threatened Species 2018.:e.T22724308A132174807.  
724 <http://dx.doi.org/10.2305/IUCN.UK.2018-2.RLTS.T22724308A132174807.en>.

725 Johnson K, Peer BD. 2001. Great-tailed grackle: *Quiscalus mexicanus*. *Birds of North America*, Incorporated.  
726 rated.

727 Johnson NK, Cicero C. 2004. New mitochondrial dna data affirm the importance of pleistocene speciation  
728 in north american birds. *Evolution*. 58(5):1122–1130.

729 Kotrschal A, Rogell B, Bundsen A, Svensson B, Zajitschek S, Brännström I, Immler S, Maklakov AA, Kolm  
730 N. 2013. Artificial selection on relative brain size in the guppy reveals costs and benefits of evolving a larger  
731 brain. *Current Biology*. 23(2):168–171.

732 Künzl C, Kaiser S, Meier E, Sachser N. 2003. Is a wild mammal kept and reared in captivity still a wild  
733 animal? *Hormones and behavior*. 43(1):187–196.

734 Landis JR, Koch GG. 1977. The measurement of observer agreement for categorical data. *biometrics*.:159–  
735 174.

736 Laumer I, Call J, Bugnyar T, Auersperg A. 2018. Spontaneous innovation of hook-bending and unbending  
737 in orangutans (*pongo abelii*). *Scientific reports*. 8(1):1–13.

738 Lefebvre L, Whittle P, Lascaris E, Finkelstein A. 1997. Feeding innovations and forebrain size in birds.  
739 *Animal Behaviour*. 53(3):549–560.

740 Lewis J. 2020. Leastcostpath: Modelling pathways and movement potential within a landscape (version  
741 1.7.4). <https://CRAN.R-project.org/package=leastcostpath>.

742 Liu X, Huang Y, Xu X, Li X, Li X, Ciais P, Lin P, Gong K, Ziegler AD, Chen A, et al. 2020. High-  
743 spatiotemporal-resolution mapping of global urban change from 1985 to 2015. *Nature Sustainability*.:1–7.

Logan C. 2016. Behavioral flexibility and problem solving in an invasive bird. *PeerJ*. 4:e1975.

Logan CJ. 2016. Behavioral flexibility in an invasive bird is independent of other behaviors. *PeerJ*. 4:e2215.

Logan CJ, Avin S, Boogert N, Buskell A, Cross FR, Currie A, Jelbert S, Lukas D, Mares R, Navarrete AF, et al. 2018. Beyond brain size: Uncovering the neural correlates of behavioral and cognitive specialization. *Comparative cognition & behavior reviews*.

Logan CJ, MacPherson M, Rowney C, Bergeron L, Seitz B, Blaisdell A, Folsom M, Johnson-Ulrich Z, McCune K. 2019.

Manrique HM, Call J. 2011. Spontaneous use of tools as straws in great apes. *Animal cognition*. 14(2):213–226.

McCune KB. 2018. Cognition gone wild: A test of the social intelligence hypothesis in wild birds [PhD thesis].

McCune KB, Jablonski P, Lee S-i, Ha RR. 2019. Captive jays exhibit reduced problem-solving performance compared to wild conspecifics. *Royal Society open science*. 6(1):181311.

McCune KB, MacPherson M, Rowney C, Bergeron L, Folsom M, Logan C. 2019.

McElreath R. 2016. Statistical rethinking: A bayesian course with examples in r and stan. CRC Press. <http://xcelab.net/rm/statistical-rethinking/>.

McElreath R. 2020. Rethinking: Statistical rethinking book package.

Mery F, Kawecki TJ. 2005. A cost of long-term memory in drosophila. *Science*. 308(5725):1148–1148.

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.

Mikhalevich I, Powell R, Logan C. 2017. Is behavioural flexibility evidence of cognitive complexity? How evolution can inform comparative cognition. *Interface Focus*. 7(3):20160121. doi:10.1098/rsfs.2016.0121. [accessed 2017 May 29]. <http://rsfs.royalsocietypublishing.org/lookup/doi/10.1098/rsfs.2016.0121>.

Pandolfino E, Deuel B, Young L. 2009. Colonization of the california’s central valley by the great-tailed grackle. *Central Valley Bird Club Bull*. 12:77–95.

Pew J, Muir PH, Wang J, Frasier TR. 2015. Related: An r package for analysing pairwise relatedness from codominant molecular markers. *Molecular Ecology Resources*. 15(3):557–561.

Post W, Poston JP, Bancroft GT. 1996. Boat-tailed grackle: *Quiscalus major*. *American Ornithologists’ Union*.

R Core Team. 2017. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org>.

Revelle W. 2017. Psych: Procedures for psychological, psychometric, and personality research. Evanston, Illinois: Northwestern University. <https://CRAN.R-project.org/package=psych>.

Rutz C, Klump BC, Komarczyk L, Leighton R, Kramer J, Wischnewski S, Sugawara S, Morrissey MB, James R, St Clair JJ, et al. 2016. Discovery of species-wide tool use in the hawaiian crow. *Nature*. 537(7620):403–407.

Selander RK, Giller DR. 1961. Analysis of sympatry of great-tailed and boat-tailed grackles. *The Condor*. 63(1):29–86.

Sevchik A, Logan CJ, Bergeron L, Blackwell A, Rowney C, Lukas D. 2019. Investigating sex differences in genetic relatedness in great-tailed grackles in tempe, arizona to infer potential sex biases in dispersal (<http://corinalogan.com/preregistrations/gdispersal.html>). In principle acceptance by PCI Ecology of the version on 29 Nov 2019 <https://github.com/corinalogan/grackles/blob/master/Files/Preregistrations/gdispersalRmd>.

786 Sol D, Duncan RP, Blackburn TM, Cassey P, Lefebvre L. 2005. Big brains, enhanced cognition, and response  
787 of birds to novel environments. *Proceedings of the National Academy of Sciences of the United States of*  
788 *America*. 102(15):5460–5465.

789 Sol D, Lefebvre L. 2000. Behavioural flexibility predicts invasion success in birds introduced to new zealand.  
790 *Oikos*. 90(3):599–605.

791 Sol D, Székely T, Liker A, Lefebvre L. 2007. Big-brained birds survive better in nature. *Proceedings of the*  
792 *Royal Society of London B: Biological Sciences*. 274(1611):763–769.

793 Sol D, Timmermans S, Lefebvre L. 2002. Behavioural flexibility and invasion success in birds. *Animal*  
794 *behaviour*. 63(3):495–502.

795 Stan Development Team. 2020. RStan: The R interface to Stan. <http://mc-stan.org/>.

796 Strimas-Mackey M, Miller E, Hochachka W. 2018. Auk: EBird data extraction and processing with awk.  
797 <https://cornelllabofornithology.github.io/auk/>.

798 Strimas-Mackey M., W.M. Hochachka, V. Ruiz-Gutierrez, O.J. Robinson, E.T. Miller, T. Auer, S. Kelling, D.  
799 Fink, Johnston A. 2016. Best practices for using eBird data. Version 1.0. ([https://cornelllabofornithology.github.io/ebird-](https://cornelllabofornithology.github.io/ebird-best-practices/)  
800 [best-practices/](https://cornelllabofornithology.github.io/ebird-best-practices/)). *Current opinion in behavioral sciences*. 12:90–96. doi:<https://doi.org/10.5281/zenodo.3620739>.

801 Taylor AH, Hunt GR, Holzhaider JC, Gray RD. 2007. Spontaneous metatool use by new caledonian crows.  
802 *Current Biology*. 17(17):1504–1507.

803 Thrasher DJ, Butcher BG, Campagna L, Webster MS, Lovette IJ. 2018. Double-digest rad sequencing  
804 outperforms microsatellite loci at assigning paternity and estimating relatedness: A proof of concept in a  
805 highly promiscuous bird. *Molecular ecology resources*. 18(5):953–965.

806 Urbanek S. 2020. RJava: Low-level r to java interface. <https://CRAN.R-project.org/package=rJava>.

807 Wehtje W. 2003. The range expansion of the great-tailed grackle (*quiscalus mexicanus* gmelin) in north  
808 america since 1880. *Journal of Biogeography*. 30(10):1593–1607.

809 Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Golemund G, Hayes A, Henry  
810 L, Hester J, et al. 2019. Welcome to the tidyverse. *Journal of Open Source Software*. 4(43):1686.  
811 doi:10.21105/joss.01686.

812 Wiens JA. 1997. Metapopulation dynamics and landscape ecology. In: *Metapopulation biology*. Elsevier.  
813 pp. 43–62.

814 Wright TF, Eberhard JR, Hobson EA, Avery ML, Russello MA. 2010. Behavioral flexibility and species  
815 invasions: The adaptive flexibility hypothesis. *Ethology Ecology & Evolution*. 22(4):393–404.

816 Wu J, Jenerette GD, Buyantuyev A, Redman CL. 2011. Quantifying spatiotemporal patterns of urbanization:  
817 The case of the two fastest growing metropolitan regions in the united states. *Ecological Complexity*. 8(1):1–  
818 8.