

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

Logan CJ<sup>1\*</sup>

McCune KB<sup>2</sup>

Chen N<sup>3</sup>

Lukas D<sup>1</sup>

16 December 2021



**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](mailto:corina_logan@eva.mpg.de)

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](https://doi.org/10.24072/pci.ecology.100062). Reviewers: Caroline Nieberding, Tim Parker, and Pizza Ka Yee Chow



**Cite as:** Logan CJ, McCune KB, 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 16 Dec 2021 <https://github.com/corinalogan/grackles/blob/0fb956040a34986902a384a1d8355de65010effd/Files/Preregistrations/gxpopbehaviorhabitat.Rmd>.

See the easy-to-read [HTML](#) version and the reproducible manuscript ([Rmd](#)) version for the code

## 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 (Chow et al., 2016; Griffin & Guez, 2014; e.g. Lefebvre et al., 1997; Sol et al., 2002, 2005, 2007; Sol & Lefebvre, 2000). However, it is an alternative non-exclusive possibility that an increase in the amount of available habitat can also facilitate a range expansion (Hanski & 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 (K. Johnson & Peer, 2001). They are behaviorally flexible (Logan, 2016a) and highly associated with human-modified environments (K. Johnson & 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. **Finally, we will 4) determine whether changes in behavioral traits facilitate the rapid GTGR expansion by comparing their behavior with BTGR on the same tests in aim 1.** 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 (Chow et al., 2016; Griffin & Guez, 2014; e.g., Lefebvre et al., 1997; Sol et al., 2002, 2005, 2007; Sol & Lefebvre, 2000). 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 (A. M. Auersperg et al., 2012; Bird & Emery, 2009; Laumer et al., 2018; Manrique & Call, 2011; e.g., Taylor et al., 2007). 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 & 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 (Goldewijk, 2001; e.g., Liu et al., 2020; Wu et al., 2011), and species associated with these habitats show differences in their behavior (Chejanovski et al., 2017; e.g., Ciani, 1986; 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 (Logan, 2016a), rapidly expanding their geographic range (Wehtje, 2003), and highly associated with human-modified environments (K. Johnson & 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 (K. Johnson & 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 relative to older populations. Specifically, we will investigate behavioral flexibility measured as reversal learning of food-filled colored tube preferences (Logan, 2016b; Logan et al., 2019), innovativeness measured as the number of loci they solve to access food on a puzzlebox (A. M. I. Auersperg et al., 2011; Logan et al., 2019), exploration measured as the latency to approach a novel object in the absence of nearby food (McCune KB et al., 2019; Mettke-Hofmann et al., 2009), 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).

Second, 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 & 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 & 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 & 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 & Giller (1961), K. Johnson & 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.

**Third, we aim to compare behavior in one population of wild-caught boat-tailed grackles with that of great-tailed grackles on the same tests in aim 1 (behavioral flexibility, innovativeness, exploration, and persistence, but not dispersal).** Similar to great-tailed grackles, boat-tailed grackles are social and polygamous, and eat human foods (Post, 1992; Post et al., 1996), which increases the ecological relevance of these tests. Determining whether great-tailed grackles perform better on these tests would provide support for the hypothesis that their behavior could be causing their rapid geographic range expansion. Alternatively, if boat-tailed and great-tailed grackles perform similarly, this would suggest that environmental, rather than behavioral, variables may play a larger role in restricting the boat-tailed grackle range expansion.

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: great-tailed grackles):** 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 & Giller (1961), K. Johnson & 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 Kotrschal 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 & 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.

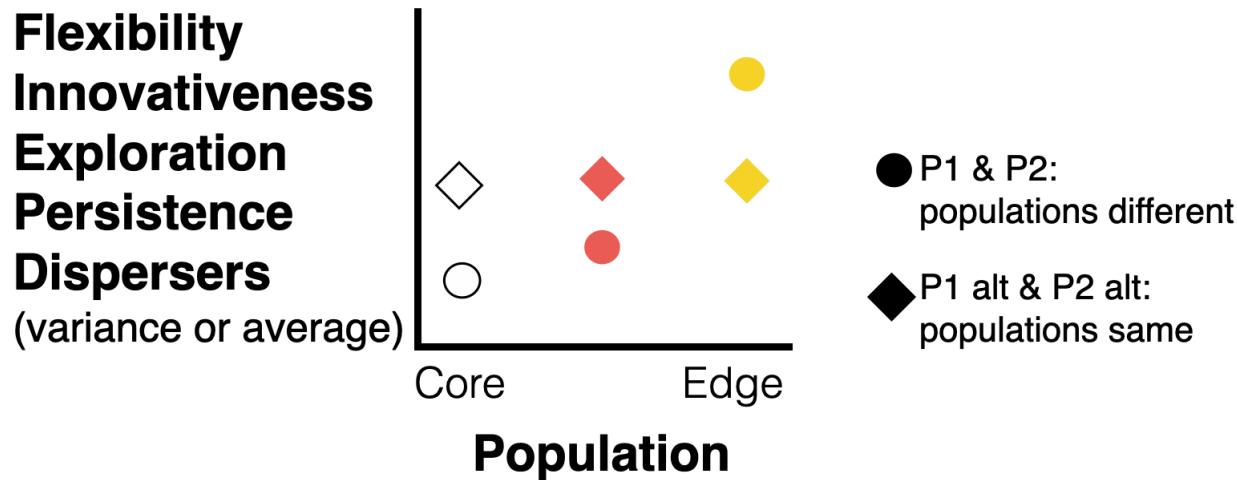


Figure 1: **Figure 1.** What is the role of behavior in a rapid range expansion? The great-tailed grackle study sites are indicated by the colored circles: edge (yellow; California), middle (red; Arizona), and core (white; Central America) and correspond with those in Figure 3.

**Q2 (dispersal behavior: great-tailed grackles):** Are there differences in dispersal behavior across the great-tailed grackle's geographic range? (Fig. 1, Table 1) **Prediction 2:** We predict

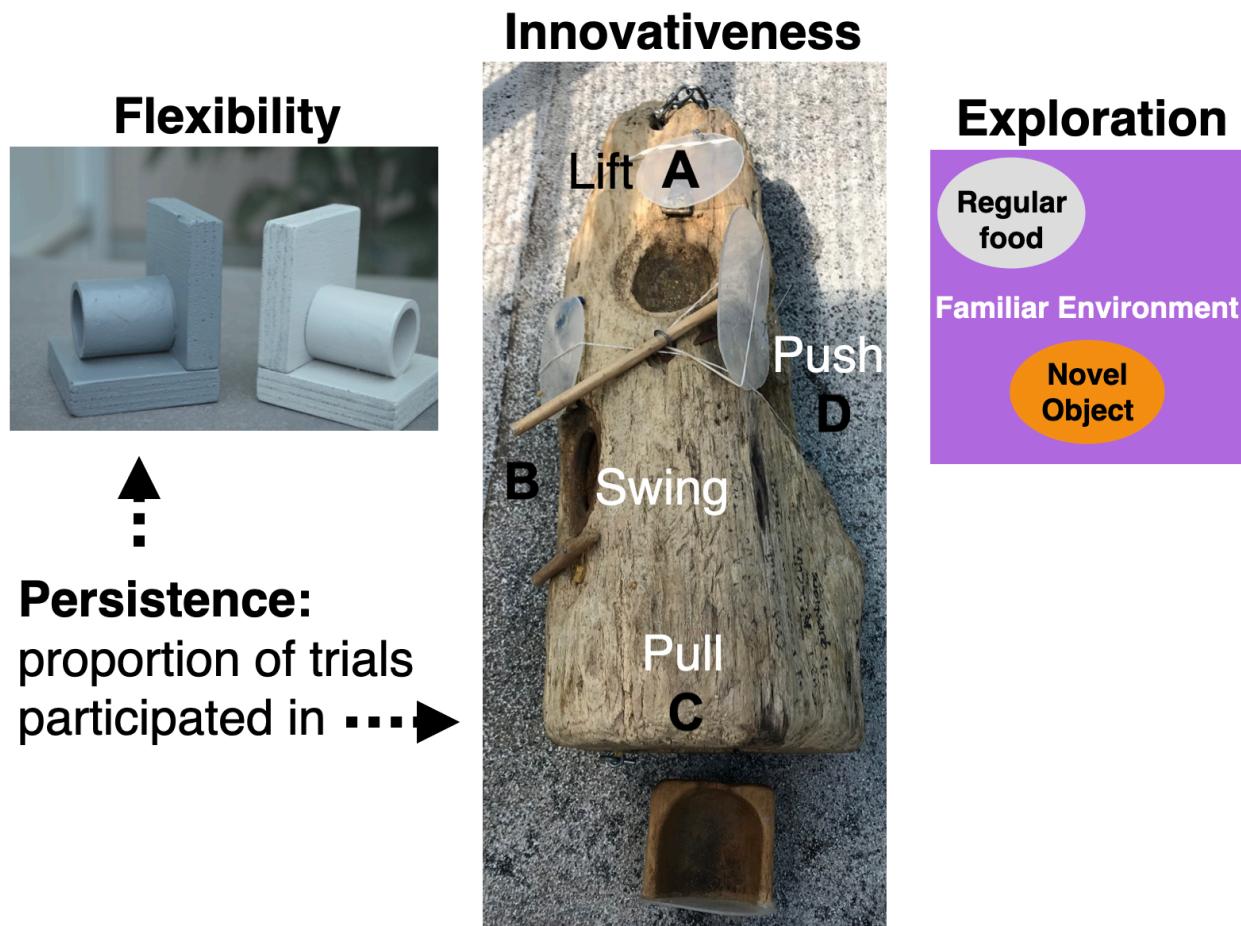


Figure 2: **Figure 2.** Experimental protocol. Great-tailed grackles from the core, middle, and edge populations will be tested for their: (top left) flexibility (number of trials to reverse a previously learned color tube-food association); (middle) innovativeness (number of options [lift, swing, pull, 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).

**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); note that this species starts breeding at age 1], and on the first year in which this species was reported (or estimated) to breed at each location (Woodland, California: Yolo Audubon Society's newsletter *The Burrowing Owl* (July 2004), which Steve Hampton shared with Logan; Tempe, Arizona: estimated based on 1945 first-sighting report in nearby Phoenix, Arizona (Wehtje, 2004) to which we added 6 years, which is the average time between first-sighting and first-breeding - see Table 3 in (Wehtje, 2003); 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 (N. K. Johnson & Cicero, 2004)).

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	1951	69	12.3	Wehtje 2003, 2004
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 & 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.

**Q4 (4 behaviors: boat-tailed grackles): Are there differences in behavioral traits (flexibility, innovation, exploration, and persistence) between boat-tailed and great-tailed grackles? (Fig 2)** **Prediction 6:** If behavior modifications are needed to adapt to new locations, then great-tailed grackles, which are rapidly expanding their geographic range (Wehtje, 2003), will have higher averages and/or larger variances than boat-tailed grackles, which are not rapidly expanding their range Wehtje (2003), in at

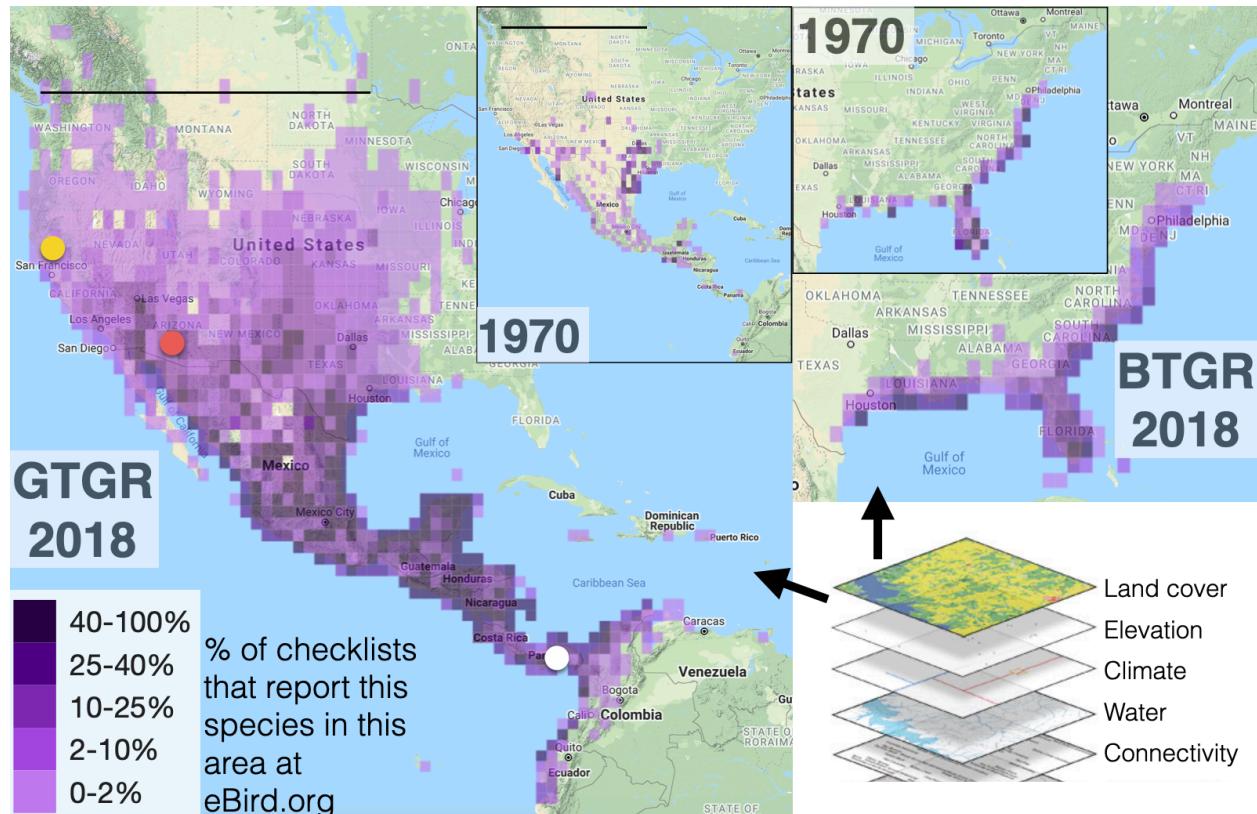
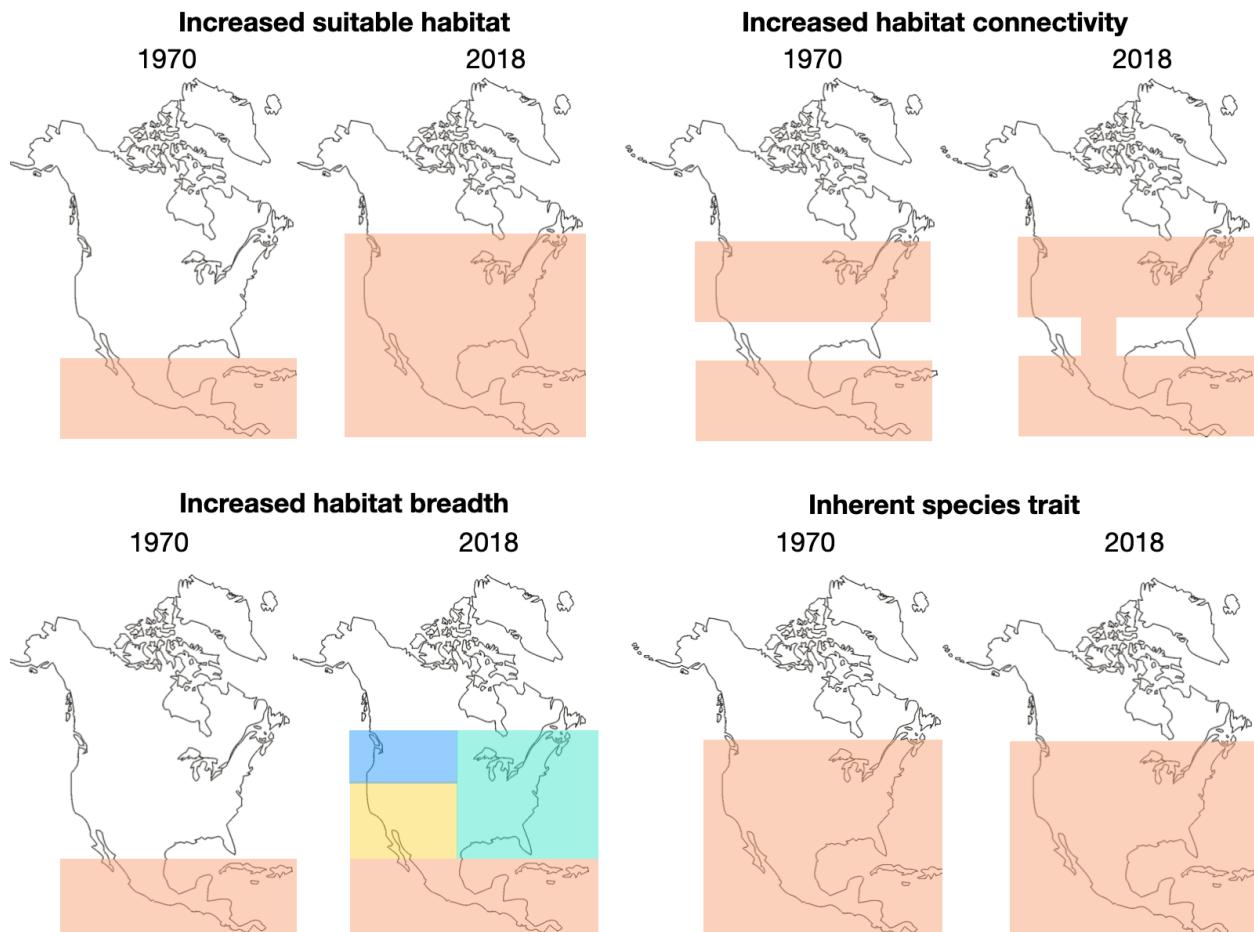


Figure 3: **Figure 3.** Ecological niche modeling of the presence/absence of great-tailed grackles (GTGR; left), which are rapidly expanding their geographic range (see inset for range change since 1970), and boat-tailed grackles (BTGR; right), which are not (see inset for range change since 1970). The 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.

least some behavioral traits (behavioral flexibility: speed at reversing a previously learned color preference, innovativeness: number of options solved on a puzzle box, exploration: latency to approach/touch a novel object, and persistence: proportion of trials participated in). Higher averages in behavioral traits indicate that each individual can exhibit more of that trait. If resources are regularly dispersed in time and space, perhaps boat-tailed grackles require less flexibility when visiting these resources and attend less to their temporal availability or the individual's food preferences. Perhaps the problems boat-tailed grackles solve do not require much exploration or persistence. Lower variances in behavioral traits indicate that there is less diversity of individuals in the population, which means that there is a lower chance that some individuals in the population would innovate new foraging techniques and be more flexible, exploratory, and persistent.

**Prediction 6 alternative 1:** Human-modified environments are suitable habitat for great-tailed and boat-tailed grackles and the amount of human-modified environments has and is increasing. If the original behaviors exhibited by these species happen to be suited to the uniformity of human-modified landscapes, then averages and/or variances of these traits will be similar in the great-tailed and boat-tailed grackles sampled. This supports the hypothesis that, because these species are closely associated with human-modified environments Wehtje (2003), which may be similar across the geographic range, 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).

**Prediction 6 alternative 2:** If boat-tailed grackles have higher averages and/or larger variances in the behavioral traits measured, this could indicate that perhaps these traits are not the primary facilitators of the great-tailed grackle's rapid geographic range expansion. Alternatively, perhaps these species differ in a life history variable that restricts the boat-tailed grackle from expanding, or there is some geographic feature that prevents the boat-tailed grackle from rapidly expanding its range (see Q3).

## 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 & 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).

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.

**Q4: Boat-tailed grackles are caught in the wild at and near Archbold Biological Station in Venus, Florida. Individuals are given colored leg bands in unique combinations for identification, brought temporarily into aviaries (the same aviaries as above) for testing, and then**

released back to the wild at their point of capture. Details about housing and feeding, etc. are the same as above for great-tailed.

**Sample size rationale** We summarize the minimum sample sizes and their associated detection limits in Table 2, which will allow us to determine whether populations are different from each other (detailed in the Analysis section for each experiment).

**Table 2.** A summary of the measure of interest in each experiment, the distribution used for the analysis, the minimum detectable difference between site means, and the minimum sample size that goes with the minimum detectable difference.

Experiment	Measurement	Distribution	Minimum difference between site means	Minimum sample size
Reversal	Phi (learning rate)	Gamma	Differences of 0.01 are likely to be detected (based on models with 20 individuals per site, however this is likely to hold for the the minimum sample size as well) (Figures 5 and 6)	15
Reversal	Lambda (random choice rate)	Gamma	Differences of 3 are likely to be detected (based on models with 20 individuals per site, however this is likely to hold for the the minimum sample size as well) (Figures 5 and 6)	15
Multiaccess box	Number of loci solved	Binomial	Differences of 1.2 loci are likely to be detected (Table 3)	15
Exploration	Latency to approach novel object	Gamma-Poisson	Differences of at least 450 seconds are likely to be detected (Table 3)	15
Persistence	Percent of trials participated in	Normal	Difference of at least 0.08 in the proportion of trials participated in (Table 3)	18

**Q1 & Q2:** We test as many great-tailed grackles 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 follow the minimum sample sizes in Table 2 with the aim that half of the grackles tested at each site are female. However, we expect to be able to test 20 grackles per site.

**Q4:** We test as many boat-tailed grackles as we can during two approximately three month field seasons, which are conducted during the non-breeding season (approximately September through March). It is time intensive to conduct the aviary test battery (1-2 months per bird at the great-tailed grackle Woodland field site), therefore we approximate that the minimum sample size per experiment will follow the minimum sample sizes in Table 2 with the aim that half of the grackles tested are female.

**Data collection stopping rule** We will stop collecting data on wild-caught great-tailed 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).

We will stop collecting data on wild-caught boat-tailed grackles in Q4 once we have reached our minimum sample size.

## Protocols and open materials

- Experimental protocols for Q1 (which are the same as for Q4) 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 A. M. I. 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 & Giller (1961), K. Johnson & 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** When the study is complete, the data will be published in the Knowledge Network for Bio-complexity's data repository (Q1, Q2, Q4).

**Randomization and counterbalancing (Q1, Q2, & Q4)** **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: LRLLRRLRLR, 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](http://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 4.1.2; 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 & 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 because watching the videos is not as clear as watching the bird participate in the trial from the aisle of the aviaries. The unweighted kappa is used when analyzing a categorical variable where the distances between the numbers are meaningless (0=incorrect choice, 1=correct choice, -1=did not participate), and the ICC is used for continuous variables where distances are meaningful (e.g., if coders disagree by a difference of 2 s rather than 5 s, this is important to account for).

**Interobserver reliability training** To pass **interobserver reliability (IOR) training**, video coders needed an ICC or Cohen's unweighted kappa score of 0.90 or greater to ensure the instructions were clear and that there was a high degree of agreement across coders. Video coders, Alexis Breen and Vincent Kiepsch, passed interobserver reliability training for exploration in a previous article (McCune KB et al., 2019) where their training results can be found.

**Lea Gihlein** (compared with experimenter's live coding):

- Reversal learning: correct choice unweighted Cohen's Kappa=1.00 (confidence boundaries=1.00-1.00, n=21 data points)
- Multiaccess box: correct choice unweighted Cohen's Kappa=1.00 (confidence boundaries=1.00-1.00, n=29 data points)
- Multiaccess box: correct choice unweighted Cohen's Kappa=1.00 (confidence boundaries=1.00-1.00, n=29 data points)

**Interobserver reliability** Interobserver reliability scores (minimum 20% of the videos: 5/? birds) were as follows:

**Lea Gihlein** (compared with experimenter's live coding):

- Reversal learning: correct choice unweighted Cohen's Kappa=1.00 (confidence boundaries=0.99-1.00, n=707 data points)
- Multiaccess box: correct choice unweighted Cohen's Kappa=0.92 (confidence boundaries=0.81-1.00, n=63 data points)
- Multiaccess box: locus solved unweighted Cohen's Kappa=1.00 (confidence boundaries=1.00-1.00, n=48 data points)

**Vincent Kiepsch** (compared with Breen's video coding):

- Exploration: latency to land on the ground unweighted Cohen's Kappa=0.998 (confidence boundaries=0.997-0.999, n=32 data points)

#### **Q1 & Q4: behavior across the great-tailed grackle range and between great-tailed and boat-tailed grackles 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. **Boat-tailed grackles will be considered a site and we will be able to examine differences between them and each of the great-tailed grackle sites.**

**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 4.1.2; R Core Team (2017)) and used the following R packages: rthinking (McElreath (2020)), rstan (Stan Development Team (2020)), and Rcpp (Eddelbuettel & François (2011)).

### Flexibility analyses Model and simulation

We modified the reversal learning Bayesian model in Blaisdell et al. (2021) to simulate and analyze population differences in reversal learning, and calculate our ability to detect differences between populations. The model accounts for every choice made in the reversal learning experiment and updates the probability of choosing either option after the choice was made depending on whether that choice contained a food reward or not. It does this by updating three main components for each choice: an attraction score, a learning rate ( $\phi$ ), and a rate of deviating from learned attractions ( $\lambda$ ).

**Equation 1 (attraction and  $\phi$ ):**  $A_{i,j,t+1} = (1 - \phi_j)A_{i,j,t} + \phi_j\pi_{i,j,t}$

Equation 1 “tells us how attractions to different behavioral options  $A_{i,j,t+1}$  (i.e., how preferable option  $i$  is to the bird  $j$  at time  $t+1$ ) change over time as a function of previous attractions  $A_{i,j,t}$  and recently experienced payoffs  $\pi_{i,j,t}$  (i.e., whether they received a reward in a given trial or not). Attraction scores thus reflect the accumulated learning history up to this point. The (bird-specific) parameter  $\phi_j$  describes the weight of recent experience. The higher the value of  $\phi_j$ , the faster the bird updates their attraction. It thus can be interpreted as the *learning or updating rate of an individual*. A value of  $\phi_j = 0.04$ , for example, means that receiving a single reward for one of the two options will shift preferences by 0.02 from initial 0.5-0.5 attractions, a value of  $\phi_j = 0.06$  will shift preferences by 0.03 and so on” (Blaisdell et al., 2021).

**Equation 2 ( $\lambda$ ):**  $P(i)_{t+1} = \frac{\exp(\lambda_j A_{i,j,t})}{\sum_{m=1}^2 \exp(\lambda_j A_{m,j,t})}$

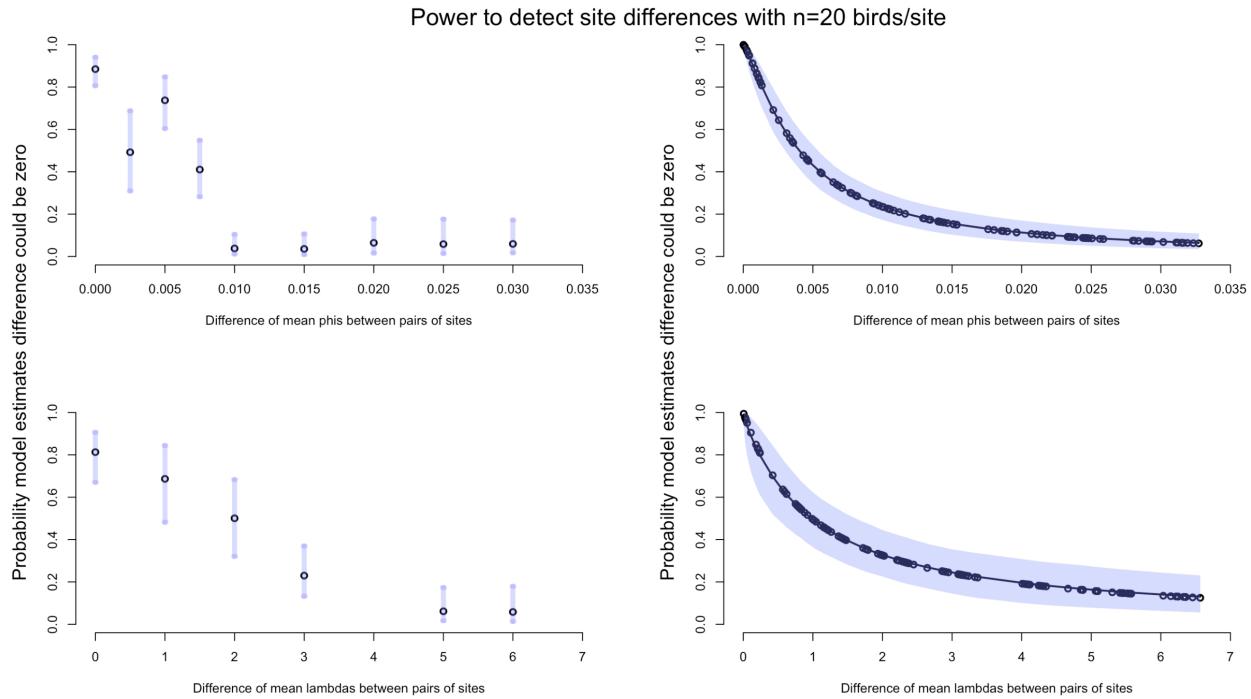
Equation 2 “expresses the probability an individual  $j$  chooses option  $i$  in the next round,  $t+1$ , based on the latent attractions. The parameter  $\lambda_j$  represents the *rate of deviating from learned attractions* of an individual

(also called inverse temperature). It controls how sensitive choices are to differences in attraction scores. As  $\lambda_j$  gets larger, choices become more deterministic, as it gets smaller, choices become more exploratory (random choice if  $\lambda_j = 0$ ). For instance, if an individual has a 0.6-0.4 preference for option A, a value of  $\lambda_j = 3$  means they choose A 65% of the time, a value of  $\lambda_j = 10$  means they choose A 88% of the time and a value of  $\lambda_j = 0.5$  means they choose A only 53% of the time” (Blaisdell et al., 2021).

As in Blaisdell et al. (2021), we, too, used previously published data on reversal learning of color tube preferences in great-tailed grackles (Logan, 2016b) to inform the model modifications. We modified the Blaisdell et al. (2021) model in a few ways: 1) we set the initial attraction score assigned to option 1 and option 2 (the light gray and dark gray tubes) to 0.1 rather than 0.0. This change assumes that there would be some inclination (rather than no inclination) for the bird to approach the tubes when they are first presented because they had been previously trained to expect food in tubes. This also allows the attraction score to decrease when a non-rewarded choice is made near the beginning of the experiment. With the previous initial attraction scores set to zero, a bird would be expected to choose the rewarded option in 100% of the trials after the first time it chose that option (attraction cannot be lower than zero, and choice is shaped by the ratio of the two attractions so that when one option is zero and the other is larger than zero, the ratio will be 100% for the rewarded option). 2) We changed the updating so that an individual would only change the attraction toward the option they chose in that trial (either decreasing their attraction toward the unrewarded option or increasing their attraction toward the rewarded option). Previously, both attractions were updated after every trial, assuming that individuals understand that the experiment is setup such that one option is always rewarded. For our birds, we instead assumed that individuals will focus on their direct experience rather than making abstract assumptions about the test. Our modification resulted in needing a higher  $\phi$  to have the same learning rate as a model where both attraction scores are updated after every trial. This change also appeared to better reflect the performance of the Santa Barbara grackles, because they had higher  $\phi$  values, which, in turn, meant lower  $\lambda$  values to reflect the performance during their initial learning. These lower  $\lambda$  values better reflect the birds’ behavior during the first reversal trials: a large  $\lambda$  value means that birds continue to choose the now unrewarded option almost 100% of the time, whereas the lower  $\lambda$  values mean that birds start to explore the rewarded option relatively soon after the switch of the rewarded option.

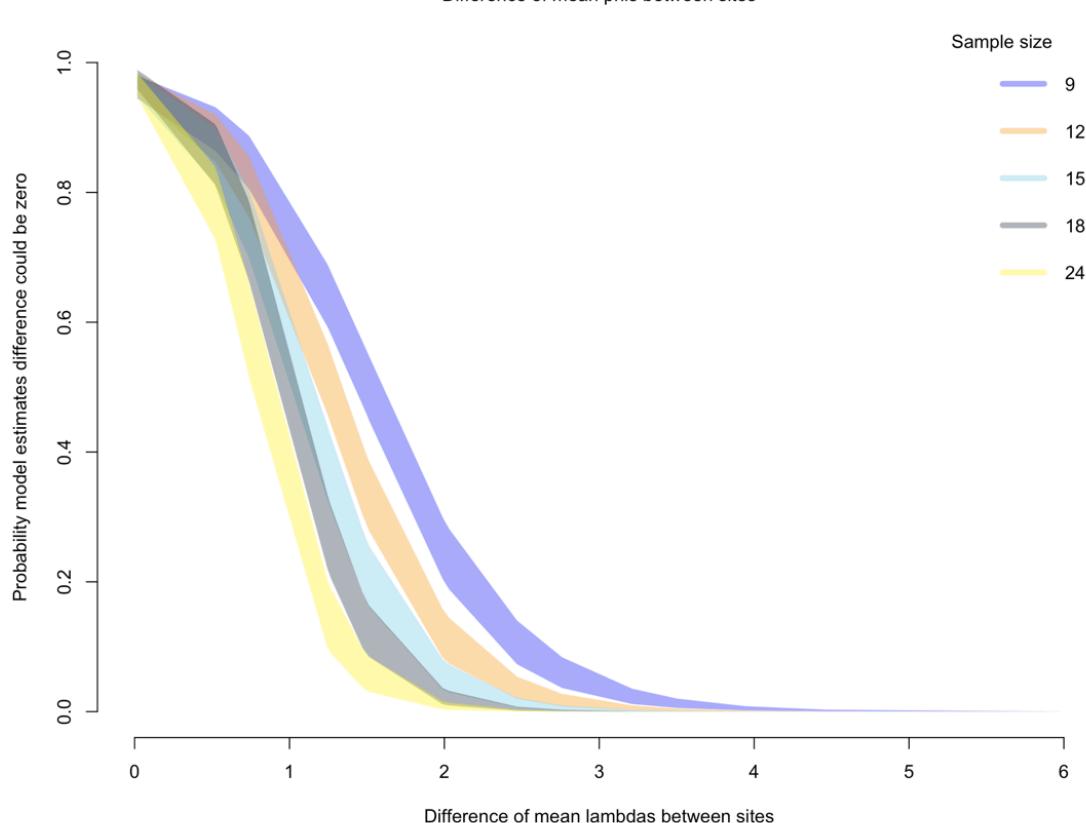
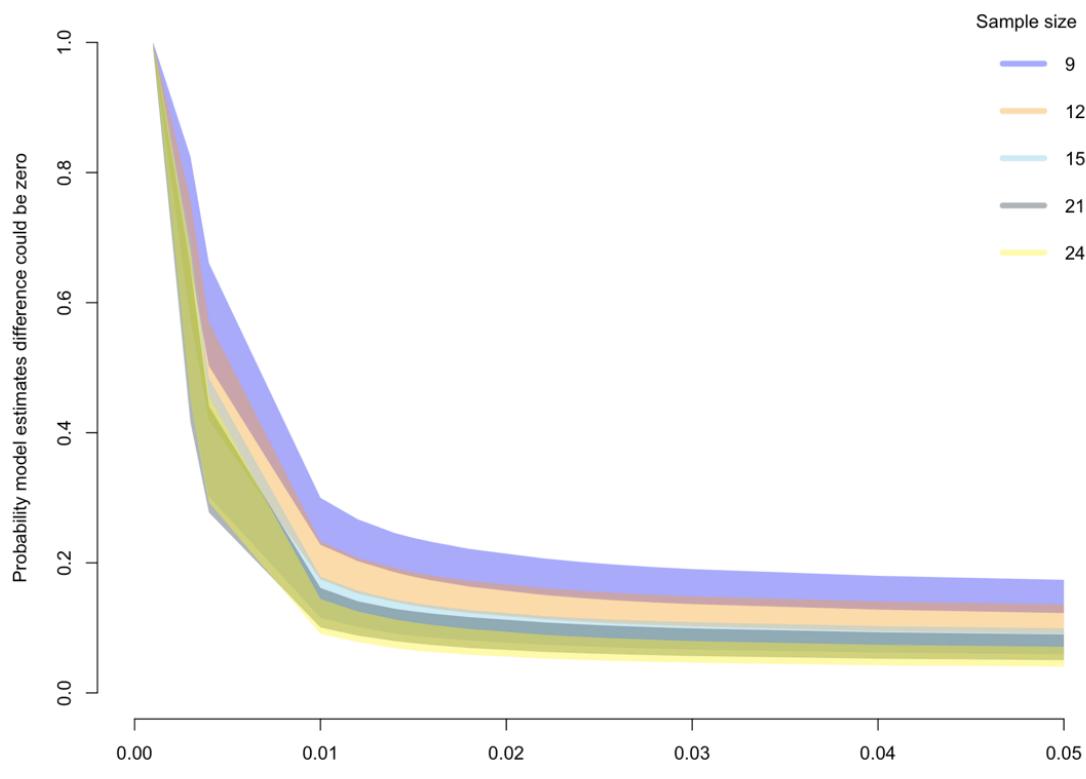
We first reanalyzed the Santa Barbara grackle data to obtain the phi and lambda values with this revised model, which informed our expectations of what a site’s mean and variance might be. Then we ran simulations, where we determined that we wanted to make the previously mentioned modifications to the stan (Team & others, 2019) model [in R, current version 4.1.2; R Core Team (2017)]. This model will be used to analyze the actual data after it is collected.

**Power analyses:** We also used the simulations to estimate our ability to detect differences in  $\phi$  and  $\lambda$  between sites based on extracting samples from the posterior distribution. We ran two different sets of simulations: the first set of simulations recreated choices for 20 birds per population across initial learning and reversal trials; the second set of simulations sampled between 9 and 24 birds from populations with pre-specified  $\phi$  and  $\lambda$  means. The first set of simulations showed that we have a very high chance of detecting that two sites are different from each other if the difference in their  $\phi$  is 0.01 or greater and/or if the difference in their  $\lambda$  is 3 or greater, based on data from 20 simulated individuals per site (Figure 5). The second set of simulations shows how different site sample sizes change detection levels: once a sample size of 15 is reached, there are only minimal differences in detection abilities compared to larger sample sizes (Figure 6). It appears that there is more variability in the  $\lambda$  estimates for each bird based on their choices, meaning that with the learning model, which estimates  $\lambda$  from the choices, the differences between sites have to be larger (than if we were able to infer lambda directly) to be reliably detected. The second set of simulations assumed that we could infer  $\lambda$  directly, however it is not possible to directly measure  $\lambda$ . Therefore, the power curves in Figure 5 are more reliable than those in Figure 6.



**Figure 5.** How small of a site difference in phi and lambda can we detect? The probability that the model estimates that the difference shown on the x axis is zero, meaning that the model assumes that it is possible that these two estimates come from a population with the same phi or lambda. Each point is the mean phi or mean lambda from one site minus the mean phi or mean lambda from another site (calculated from 20 individuals per site) for all pairwise comparisons for all 32 sites (for a total of 496 pairwise comparisons). Left panels: error bars=89% compatibility intervals. Right panels: shaded areas=97% prediction intervals.

### Power to detect site differences



**Figure 6.** How do detection differences vary according to sample size differences? The probability that the model estimates that the difference shown on the x axis is zero, meaning that the model assumes that it is possible that these two estimates come from a population with the same phi or lambda. The x-axis is the mean phi or mean lambda from one site minus the mean phi or mean lambda from another site for all pairwise comparisons for all 14 sites (for a total of 91 pairwise comparisons). Each shaded region is the 97% prediction interval for that particular sample size.

### 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), McCune et al. (2019)).

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

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

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

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

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

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 3). 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 3). Note: the Arizona sample size is 11 for the multiaccess log and 17 on a similar multiaccess box.

**Table 3.** Simulation outputs from varying sample size (n), and  $\alpha$  means and standard deviations. We calculate pairwise contrasts between the estimated means 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.



Response variable	n	mu1	mu2	mu3	mu1_sd	mu2_sd	mu3_sd	Difference S1-S2	Difference S1-S3	Difference S2-S3	Notes
loci solved	60	1.90	2.10	3.60	0.50	0.50	0.50	No	No	No	No
loci solved	60	1.90	2.10	2.20	0.50	0.50	0.50	Yes	Yes	Yes	Yes
loci solved	60	1.90	2.10	2.30	0.50	0.50	0.50	Yes	Yes	Yes	Yes
loci solved	60	1.90	2.10	3.50	0.50	0.50	0.50	No	No	No	No
loci solved	60	1.90	2.10	3.00	0.50	0.50	0.50	Yes	No	No	No
loci solved	60	1.90	2.10	2.80	0.50	0.50	0.50	Yes	No	Yes	No
loci solved	60	1.80	2.10	3.00	0.50	0.50	0.50	Yes	Yes	No	No
loci solved	60	2.00	2.50	3.00	0.50	0.50	0.50	No	Yes	No	No
loci solved	60	2.00	2.50	3.10	0.50	0.50	0.50	No	No	Yes	
loci solved	60	1.90	2.50	3.20	0.50	0.50	0.50	Yes	No	No	No
loci solved	60	1.80	2.50	3.30	0.50	0.50	0.50	No	No	Yes	
loci solved	60	1.70	2.50	3.40	0.50	0.50	0.50	No	No	No	No
loci solved	60	1.70	2.50	3.40	1.00	1.00	1.00	No	No	No	No
loci solved	60	1.70	2.50	3.40	1.50	1.50	1.50	Yes	No	No	No
loci solved	60	1.70	2.50	3.40	1.30	1.30	1.30	Yes	Yes	Yes	Yes
loci solved	60	1.00	2.00	3.00	0.50	0.50	0.50	No	No	Yes	
loci solved	60	1.00	2.00	3.00	0.50	0.50	0.50	No	No	No	No
loci solved	60	1.00	2.00	3.00	0.30	0.40	0.50	No	No	No	No
loci solved	60	1.00	2.00	3.00	0.60	0.70	0.50	No	No	No	No
loci solved	60	1.00	2.00	3.00	0.70	0.70	0.70	No	No	No	No
loci solved	60	1.00	2.00	3.00	0.90	0.90	0.90	No	No	No	No
loci solved	60	1.00	2.00	3.00	1.00	1.00	1.00	No	No	No	No
loci solved	60	1.00	2.00	3.00	1.50	1.50	1.50	Yes	No	No	No
loci solved	60	1.00	2.00	3.00	1.30	1.50	1.50	Yes	No	No	No
loci solved	60	1.00	2.00	3.00	1.10	1.50	1.50	No	No	No	No
loci solved	60	1.00	2.00	3.00	1.20	1.50	1.50	No	No	No	No
loci solved	45	1.00	2.00	3.00	0.50	0.50	0.50	Yes	No	No	No
loci solved	45	0.90	2.00	3.10	0.50	0.50	0.50	No	No	Yes	
loci solved	45	0.80	2.00	3.20	0.50	0.50	0.50	No	No	No	No
loci solved	45	0.80	2.00	3.20	1.00	1.00	1.00	Yes	No	No	No
latency	45	5.70	6.90	7.60	1000.00	1000.00	1000.00	No	No	No	No
latency	45	5.80	6.90	7.50	1000.00	1000.00	1000.00	No	No	No	No

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:

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

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

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:

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

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

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.

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 3). 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  $\sim \text{Binomial}(\text{totaltrials}, p)$  *[likelihood]*

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

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

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

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

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 3).

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 ((J. D. 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 ( $V=1$ ,  $nu=0$ ) (J. Hadfield, 2014). We will ensure the GLMM shows acceptable convergence (i.e., lag time autocorrelation values  $<0.01$ ; (J. D. Hadfield, 2010)), and adjust parameters if necessary.

## 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.

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

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

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

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

**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, mangrove) - 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 & 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 & 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 & 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 & 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 climactic 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 M. 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` (M. Strimas-Mackey et al. (2018)), `dismo` (Hijmans et al. (2017)), `raster` (Hijmans (2020)), `maptools` (Bivand & Lewin-Koh (2019)), `tidyverse` (Wickham et al. (2019)), `rgdal` (Bivand et al. (2019)), `rJava` (Urbanek (2020)), and `elevatr` (Hollister & 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.

## 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.

**Marfori:** Data collection (Q1), revising/editing.

**LeGrande-Rolls:** Data collection (Q1 & Q2), revising/editing.

**Summers (Q3):** Data collection, ecological niche modeling, data 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; Richard McElreath for project support; and Aaron Blackwell and Ken Kosik for being the UCSB sponsors of the Cooperation Agreement with the Max Planck Institute for Evolutionary Anthropology.

## K. REFERENCES

- Auersperg, A. M. I., Bayern, A. M. P. von, Gajdon, G. K., 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. <https://doi.org/10.1371/journal.pone.0020231>
- Auersperg, A. M., Szabo, B., Von Bayern, A. M., & Kacelnik, A. (2012). Spontaneous innovation in tool manufacture and use in a goffin's cockatoo. *Current Biology*, 22(21), R903–R904.
- Bird, C. D., & Emery, N. J. (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>
- Blaisdell, A., Seitz, B., Rowney, C., Folsom, M., MacPherson, M., Deffner, D., & Logan, C. J. (2021). *Do the more flexible individuals rely more on causal cognition? Observation versus intervention in causal inference in great-tailed grackles (version 5 of this preprint has been peer reviewed and recommended by peer community in ecology [https://doi.org/10.24072/pci.ecology.100076])*. <https://doi.org/10.31234/osf.io/z4p6s>
- Bolker, B. M. (2008). *Ecological models and data in r*. Princeton University Press.
- Carpenter, B., Gelman, A., Hoffman, M. D., 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, Z. A., Avilés-Rodríguez, K. J., Lapiedra, O., Preisser, E. L., & Kolbe, J. J. (2017). An experimental evaluation of foraging decisions in urban and natural forest populations of anolis lizards. *Urban Ecosystems*, 20(5), 1011–1018.
- Chow, P. K. Y., Lea, S. E., & Leaver, L. A. (2016). How practice makes perfect: The role of persistence, flexibility and learning in problem-solving efficiency. *Animal Behaviour*, 112, 273–283. <https://doi.org/10.1016/j.anbehav.2015.11.014>
- Ciani, A. C. (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, E. C., & Collias, N. E. (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. <https://doi.org/10.18637/jss.v040.i08>

- Fedderspiel, I. G., Garland, A., Guez, D., Bugnyar, T., Healy, S. D., Güntürkün, O., & Griffin, A. S. (2017). Adjusting foraging strategies: A comparison of rural and urban common mynas (*acridotheres tristis*). *Animal Cognition*, 20(1), 65–74.
- Gamer, M., Lemon, J., Gamer, M. M., Robinson, A., & Kendall's, W. (2012). Package ‘irr.’ *Various Coefficients of Interrater Reliability and Agreement*.
- Goldewijk, K. K. (2001). Estimating global land use change over the past 300 years: The HYDE database. *Global Biogeochemical Cycles*, 15(2), 417–433.
- Griffin, A. S., & Guez, D. (2014). Innovation and problem solving: A review of common mechanisms. *Behavioural Processes*, 109, 121–134. <https://doi.org/10.1016/j.beproc.2014.08.027>
- Hadfield, J. (2014). *MCMCglmm course notes*. <http://cran.r-project.org/web/packages/MCMCglmm/vignettes/CourseNotes.pdf>
- Hadfield, J. D. (2010). MCMC methods for multi-response generalized linear mixed models: The MCMCglmm R package. *Journal of Statistical Software*, 33(2), 1–22. <http://www.jstatsoft.org/v33/i02/>
- Hanski, I., & Gilpin, M. (1991). Metapopulation dynamics: Brief history and conceptual domain. *Biological Journal of the Linnean Society*, 42(1-2), 3–16.
- Hijmans, R. J. (2020). *Raster: Geographic data analysis and modeling*. <https://CRAN.R-project.org/package=raster>
- Hijmans, R. J., Phillips, S., Leathwick, J., & Elith, J. (2017). *Dismo: Species distribution modeling*. <https://CRAN.R-project.org/package=dismo>
- Hollister, J., & Tarak Shah. (2017). *Elevatr: Access elevation data from various APIs*. <http://github.com/usepa/elevatr>
- Hutcheon, J. A., Chiolero, A., & Hanley, J. A. (2010). Random measurement error and regression dilution bias. *Bmj*, 340, c2289. <https://doi.org/10.1136/bmj.c2289>
- International, B. (2018). *Quiscalus mexicanus*. *The IUCN Red List of Threatened Species 2018*, e.T22724308A132174807. <http://dx.doi.org/10.2305/IUCN.UK.2018-2.RLTS.T22724308A132174807.en>
- Johnson, K., & Peer, B. D. (2001). *Great-tailed grackle: Quiscalus mexicanus*. Birds of North America, Incorporated.
- Johnson, N. K., & Cicero, C. (2004). New mitochondrial DNA data affirm the importance of pleistocene speciation in north american birds. *Evolution*, 58(5), 1122–1130.
- Kotrschal, A., Rogell, B., Bundsen, A., Svensson, B., Zajitschek, S., Brännström, I., Immler, S., Maklakov, A. A., & Kolm, N. (2013). Artificial selection on relative brain size in the guppy reveals costs and benefits of evolving a larger brain. *Current Biology*, 23(2), 168–171.
- Künzl, C., Kaiser, S., Meier, E., & Sachser, N. (2003). Is a wild mammal kept and reared in captivity still a wild animal? *Hormones and Behavior*, 43(1), 187–196.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159–174.
- Laumer, I., Call, J., Bugnyar, T., & Auersperg, A. (2018). Spontaneous innovation of hook-bending and unbending in orangutans (*pongo abelii*). *Scientific Reports*, 8(1), 1–13.
- Lefebvre, L., Whittle, P., Lascaris, E., & Finkelstein, A. (1997). Feeding innovations and forebrain size in birds. *Animal Behaviour*, 53(3), 549–560. <https://doi.org/10.1006/anbe.1996.0330>
- Lewis, J. (2020). *Leastcostpath: Modelling pathways and movement potential within a landscape (version 1.7.4)*. <https://CRAN.R-project.org/package=leastcostpath>
- Liu, X., Huang, Y., Xu, X., Li, X., Li, X., Ciais, P., Lin, P., Gong, K., Ziegler, A. D., Chen, A., & others. (2020). High-spatiotemporal-resolution mapping of global urban change from 1985 to 2015. *Nature Sustainability*, 1–7.

- Logan, C. J. (2016a). Behavioral flexibility and problem solving in an invasive bird. *PeerJ*, 4, e1975. <https://doi.org/10.7717/peerj.1975>
- Logan, C. J. (2016b). Behavioral flexibility in an invasive bird is independent of other behaviors. *PeerJ*, 4, e2215.
- Logan, C. J., Avin, S., Boogert, N., Buskell, A., Cross, F. R., Currie, A., Jelbert, S., Lukas, D., Mares, R., Navarrete, A. F., & others. (2018). Beyond brain size: Uncovering the neural correlates of behavioral and cognitive specialization. *Comparative Cognition & Behavior Reviews*.
- Logan, C. J., MacPherson, M., Rowney, C., Bergeron, L., Seitz, B., Blaisdell, A., Folsom, M., Johnson-Ulrich, Z., & McCune, K. B. (2019). Is behavioral flexibility manipulatable and, if so, does it improve flexibility and problem solving in a new context? *In Principle Acceptance by PCI Ecology of the Version on 26 Mar 2019*.
- Manrique, H. M., & Call, J. (2011). Spontaneous use of tools as straws in great apes. *Animal Cognition*, 14(2), 213–226.
- McCune, K. B. (2018). *Cognition gone wild: A test of the social intelligence hypothesis in wild birds* [PhD thesis].
- McCune, K. B., Jablonski, P., Lee, S., & Ha, R. R. (2019). Captive jays exhibit reduced problem-solving performance compared to wild conspecifics. *Royal Society Open Science*, 6(1), 181311.
- McCune, K. B., Jablonski, P., Lee, S., & Ha, R. R. (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). Is behavioral flexibility linked with exploration, but not boldness, persistence, or motor diversity? *In Principle Acceptance by PCI Ecology of the Version on 27 Mar 2019*. [http://corinalogan.com/Preregistrations/g\\_exploration.html](http://corinalogan.com/Preregistrations/g_exploration.html)
- McElreath, R. (2016). *Statistical rethinking: A bayesian course with examples in r and stan*. CRC Press. <https://doi.org/10.1201/9781315372495>
- McElreath, R. (2020). *Rethinking: Statistical rethinking book package*.
- Mery, F., & Kawecki, T. J. (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. <https://doi.org/10.1098/rsfs.2016.0121>
- Pew, J., Muir, P. H., Wang, J., & Frasier, T. R. (2015). Related: An r package for analysing pairwise relatedness from codominant molecular markers. *Molecular Ecology Resources*, 15(3), 557–561.
- Post, W. (1992). Dominance and mating success in male boat-tailed grackles. *Animal Behaviour*, 44(5), 917–929.
- Post, W., Poston, J. P., & Bancroft, G. T. (1996). *Boat-tailed grackle: Quiscalus major*. American Ornithologists' Union.
- R Core Team. (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org>
- Revelle, W. (2017). *Psych: Procedures for psychological, psychometric, and personality research*. Northwestern University. <https://CRAN.R-project.org/package=psych>
- Rutz, C., Klump, B. C., Komarzyk, L., Leighton, R., Kramer, J., Wischnewski, S., Sugarsawa, S., Morrissey, M. B., James, R., St Clair, J. J., & others. (2016). Discovery of species-wide tool use in the hawaiian crow. *Nature*, 537(7620), 403–407.

- Selander, R. K., & Giller, D. R. (1961). Analysis of sympatry of great-tailed and boat-tailed grackles. *The Condor*, 63(1), 29–86.
- Sevchik, A., Logan, C.J., 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. In *Principle Acceptance by PCI Ecology of the Version on 29 Nov 2019*. <http://corinalogan.com/Preregistrations/gdispersal.html>
- Sol, D., Duncan, R. P., Blackburn, T. M., Cassey, P., & Lefebvre, L. (2005). Big brains, enhanced cognition, and response of birds to novel environments. *Proceedings of the National Academy of Sciences of the United States of America*, 102(15), 5460–5465. <https://doi.org/10.1073/pnas.0408145102>
- Sol, D., & Lefebvre, L. (2000). Behavioural flexibility predicts invasion success in birds introduced to new zealand. *Oikos*, 90(3), 599–605. <https://doi.org/10.1034/j.1600-0706.2000.900317.x>
- Sol, D., Székely, T., Liker, A., & Lefebvre, L. (2007). Big-brained birds survive better in nature. *Proceedings of the Royal Society of London B: Biological Sciences*, 274(1611), 763–769.
- Sol, D., Timmermans, S., & Lefebvre, L. (2002). Behavioural flexibility and invasion success in birds. *Animal Behaviour*, 63(3), 495–502.
- Stan Development Team. (2020). *RStan: The R interface to Stan*. <http://mc-stan.org/>
- Strimas-Mackey, M., Miller, E., & Hochachka, W. (2018). *Auk: eBird data extraction and processing with AWK*. <https://cornelllabornithology.github.io/auk/>
- Strimas-Mackey, M., W. M. Hochachka, V. Ruiz-Gutierrez, O. J. Robinson, E. T. Miller, T. Auer, S. Kelling, D. Fink, & Johnston, A. (2016). Best practices for using eBird data. Version 1.0. (<Https://cornelllabornithology.github.io/ebird-best-practices/>). *Current Opinion in Behavioral Sciences*, 12, 90–96. <Https://doi.org/10.5281/zenodo.3620739>
- Taylor, A. H., Hunt, G. R., Holzhaider, J. C., & Gray, R. D. (2007). Spontaneous metatool use by new caledonian crows. *Current Biology*, 17(17), 1504–1507.
- Team, S. D., & others. (2019). Stan modeling language users guide and reference manual. *Technical Report*.
- Thrasher, D. J., Butcher, B. G., Campagna, L., Webster, M. S., & Lovette, I. J. (2018). Double-digest RAD sequencing outperforms microsatellite loci at assigning paternity and estimating relatedness: A proof of concept in a highly promiscuous bird. *Molecular Ecology Resources*, 18(5), 953–965.
- Urbanek, S. (2020). *rJava: Low-level r to java interface*. <Https://CRAN.R-project.org/package=rJava>
- Wehtje, W. (2003). The range expansion of the great-tailed grackle (*quiscalus mexicanus gmelin*) in north america since 1880. *Journal of Biogeography*, 30(10), 1593–1607. <Https://doi.org/10.1046/j.1365-2699.2003.00970.x>
- Wehtje, W. (2004). *The great-tailed grackle (quiscalus mexicanus gmelin) in the western united states of america: Range expansion and secondary contact between subspecies*. University of California, Riverside.
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., ... Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686. <Https://doi.org/10.21105/joss.01686>
- Wiens, J. A. (1997). Metapopulation dynamics and landscape ecology. In *Metapopulation biology* (pp. 43–62). Elsevier.
- Wright, T. F., Eberhard, J. R., Hobson, E. A., Avery, M. L., & Russello, M. A. (2010). Behavioral flexibility and species invasions: The adaptive flexibility hypothesis. *Ethology Ecology & Evolution*, 22(4), 393–404.
- Wu, J., Jenerette, G. D., Buyantuyev, A., & Redman, C. L. (2011). Quantifying spatiotemporal patterns of urbanization: The case of the two fastest growing metropolitan regions in the united states. *Ecological Complexity*, 8(1), 1–8.