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

Logan CJ[1]* McCune KB[2] LeGrande-Rolls C[1] Marfori Z[1] Hubbard J[3]
Lukas D[1]

2023-07-28

Open… access code peer review data

Affiliations: 1) Max Planck Institute for Evolutionary Anthropology, 2) University of California Santa Barbara, 3) Animal Behavior Graduate Group, University of California Davis. *Corresponding author: corina_logan@eva.mpg.de

This is the post-study manuscript of the preregistration that was pre-study peer reviewed and received an In Principle Recommendation on 16 Dec 2021 by:

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

Preregistration: html, pdf, rmd

Post-study manuscript (submitted to PCI Ecology for post-study peer review on 12 April 2023, revised and resubmitted in June 2023): preprint pdf at EcoEvoRxiv, rmd

ABSTRACT

It is generally thought that behavioral flexibility, the ability to change behavior when circumstances change, plays an important role in the ability of species to rapidly expand their geographic range. Great-tailed grackles (*Quiscalus mexicanus*) are a social, polygamous species that is rapidly expanding its geographic range by settling in new areas and habitats. They are behaviorally flexible and highly associated with human-modified environments, eating a variety of human foods in addition to foraging on insects and on the ground for other natural food items. They offer an opportunity to assess the role of behavior change over the course of their expansion. We compared behavior in wild-caught grackles from two populations across their range (an older population in the middle of the northern expansion front: Tempe, Arizona, and a more recent population on the northern edge of the expansion front: Woodland, California) to investigate whether certain behaviors (flexibility, innovativeness, exploration, and persistence) have higher averages and variances in the newer or older population. We found that grackles in the edge population had a higher flexibility variance (measured by reversal learning) and a higher persistence average (they participated in a larger proportion of trials), and that there were no population differences in average levels of flexibility, innovativeness (number of loci solved on a multiaccess box), or exploration (latency to approach a novel environment). Our results elucidated that individuals differentially expressing a particular behavior in an edge population could facilitate the rapid geographic range expansion of great-tailed grackles, and we found no support for the importance of several traits that were hypothesized to be involved in such an expansion. Our findings highlight the value of population studies and of breaking down cognitive concepts into direct measures of individual abilities to better understand how species might adapt to novel circumstances.

INTRODUCTION

It is generally thought that behavioral flexibility (hereafter, “flexibility”) plays an important role in the ability of a species to rapidly expand their geographic range (e.g., Lefebvre et al., 1997; Sol & Lefebvre, 2000; Sol et al., 2002, 2005, 2007; Griffin & Guez, 2014; Chow et al., 2016). It is predicted that flexibility, the ability to change behavior when circumstances change through packaging information and making it available to other cognitive processes (see Mikhalevich et al., 2017 for theoretical background on our flexibility definition), as well as exploration (latency to explore a novel environment or object) and innovation (creating new behaviors or using existing behaviors in a new context, Griffin & Guez, 2014) facilitate the expansion of individuals into completely new areas. However, the role of these behaviors in the process of establishing a population in a particular area is predicted to diminish after a certain number of generations (Wright et al., 2010). In support of this, experimental studies have shown that latent abilities are primarily expressed in a time of need (e.g., Taylor et al., 2007; Bird & Emery, 2009; Manrique & Call, 2011; Auersperg et al., 2012; Laumer et al., 2018).

To determine whether a behavior (e.g., flexibility, innovativeness, exploration, persistence) is involved in a rapid geographic range expansion, direct measures of behaviors in individuals 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). Flexibility, the ability to recognize that something about the environment has changed and decide to consider other options for deploying behavior (Mikhalevich et al., 2017), is distinct from innovation, which is the specific stringing together of particular actions in a new way or in a new context (Griffin & Guez, 2014). Innovative behavior can be related to flexibility in that it may occur in response to the decision to change behavior in some way. Investigations of behavior in invasive species and species that are rapidly expanding their geographic ranges that compare edge versus core populations are rare. Behavioral evidence from invasive species indicates that Common mynas (*Sturnus tristis*) on the invasion front are more innovative than those from populations away from the front as well as those in their native range (Cohen et al., 2020). Similarly, spiders (*Cyrtophora citricola*) and bank voles (*Myodes glareolus*) from edge populations are less exploratory than those from core populations (Chuang & Riechert, 2021; Eccard et al., 2022). An increase in innovation in newly established populations could facilitate new foraging techniques and the ability to exploit new food sources (Griffin et al., 2016), while a decrease in exploration could reduce their risk of encountering danger in a new area. More data from more species is required to discover whether these results are generalizable to an invasion or rapid range expansion context. As such, it is important to decide which measures are the best proxies of the behavior in question. For example, exploration is often measured as activity levels (e.g., Fox et al., 2009; Logan, 2016a), however it is important to distinguish activity levels, which could be an indicator of stress, from the curiosity to investigate novelty (Mettke-Hofmann et al., 2002). The latter can be accomplished by placing a novel environment or object inside of the familiar environment, thus making it optional to approach the novel element. Additionally, we can distinguish exploration from boldness through variation in food deprivation or placement of food. For boldness, the behavioral response to a potential threat, subjects are usually food deprived and then a preferred food item is placed next to the novel object (Réale et al., 2007). Whereas, in exploration assays, the regular maintenance diet is provided far away from the novel element to assess the willingness to investigate novelty without the need for food (Mettke-Hofmann et al., 2002). The latter ensures that the individual approaches the novel element primarily because they are internally motivated to explore something new.

Persistence behavior, "a measure of task-directed motivation" (Griffin & Guez 2014), to our knowledge, has not been investigated across populations of species that are rapidly expanding their geographic ranges. However, it could facilitate a range expansion through improving problem solving success (Morand-Ferron et al., 2011) and efficiency (Chow et al., 2016). There is some indication that this could be the case in a cross-species comparison of Invasive mynas who were found to be more persistent than native noisy miners (*Manorina melanocephala*) even though both species are successful in urban environments (Griffin & Diquelou, 2015). Persistence is measured in a variety of ways (e.g., work time, number of touches to the test apparatus, number/frequency of unsuccessful manipulations, etc., see Griffin & Guez, 2014 for a review), which makes it a difficult variable to compare across studies. Many measures of persistence are resource intensive to collect because they involve hundreds of hours of video coding, which could prohibit some researchers from being able to measure this variable due to time and financial constraints. Therefore, we developed an easy to calculate measure that we believe better represents task-directed motivation in grackles: the number of trials participated in divided by the total number of trials offered.

We expect that the actual act of continuing a range expansion relies on flexibility, exploration, innovation, and persistence. It is therefore likely that these behaviors are expressed more on the edge of the expansion range where there have not been many generations to accumulate relevant knowledge about or genetic adaptations to the environment. Our study aims to test whether behavioral flexibility, innovativeness, exploration, and persistence play a role in the rapid geographic range expansion of great-tailed grackles (*Quiscalus mexicanus*). Great-tailed grackles are behaviorally flexible (Logan, 2016b), rapidly expanding their geographic range (Wehtje, 2003), and highly associated with human-modified environments (Johnson & Peer, 2001), thus offering an opportunity to assess the role of behavior across their expansion. This social, polygamous species eats a variety of human foods in addition to foraging on insects and on the ground for other natural food items (Johnson & Peer, 2001). This opportunistic foraging behavior increases the ecological relevance of comparative cognition experiments that measure individual behavior abilities: grackles eat at outdoor cafes, from garbage cans, and on crops, where 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). Consequently, tests involving human-made apparatuses are ecologically relevant for this species. We compared behavior in wild-caught great-tailed grackles from two populations across their range. We use previously published data from Logan et al. (2023a) for an older population in the middle of the northern expansion front in Tempe, Arizona, as well as new data collected on a more recent population on the northern edge of the expansion front in Woodland, California (Figure 1, Table 1). We investigated whether certain behaviors had higher averages and variances in the edge population relative to the older population. Specifically, we investigated behavioral flexibility, measured as reversal learning of food-filled colored tube preferences (Logan, 2016a; Logan et al., 2023a); innovativeness, measured as the number of loci they solve to access food from a puzzle box (Auersperg et al., 2011; Logan et al., 2023a); exploration, measured as the latency in seconds to approach a novel environment in the absence of nearby food (Mettke-Hofmann et al., 2009; McCune et al., 2019b); and persistence, measured as the proportion of trials they participated in during the flexibility and innovativeness experiments (Figure 2). While it is possible for individuals in the wild to learn asocially and socially about new foods or foraging techniques to assess whether the risks are low enough to encourage exploration behavior, we focused on measuring these four behaviors in an asocial context to allow us to obtain the individual's actual cognitive performance (i.e., in the absence of dominant individuals who might hinder subordinates from participating). There could be multiple mechanisms underpinning the results, however our aim was to narrow down the role of changes in behavior in the range expansion of great-tailed grackles.

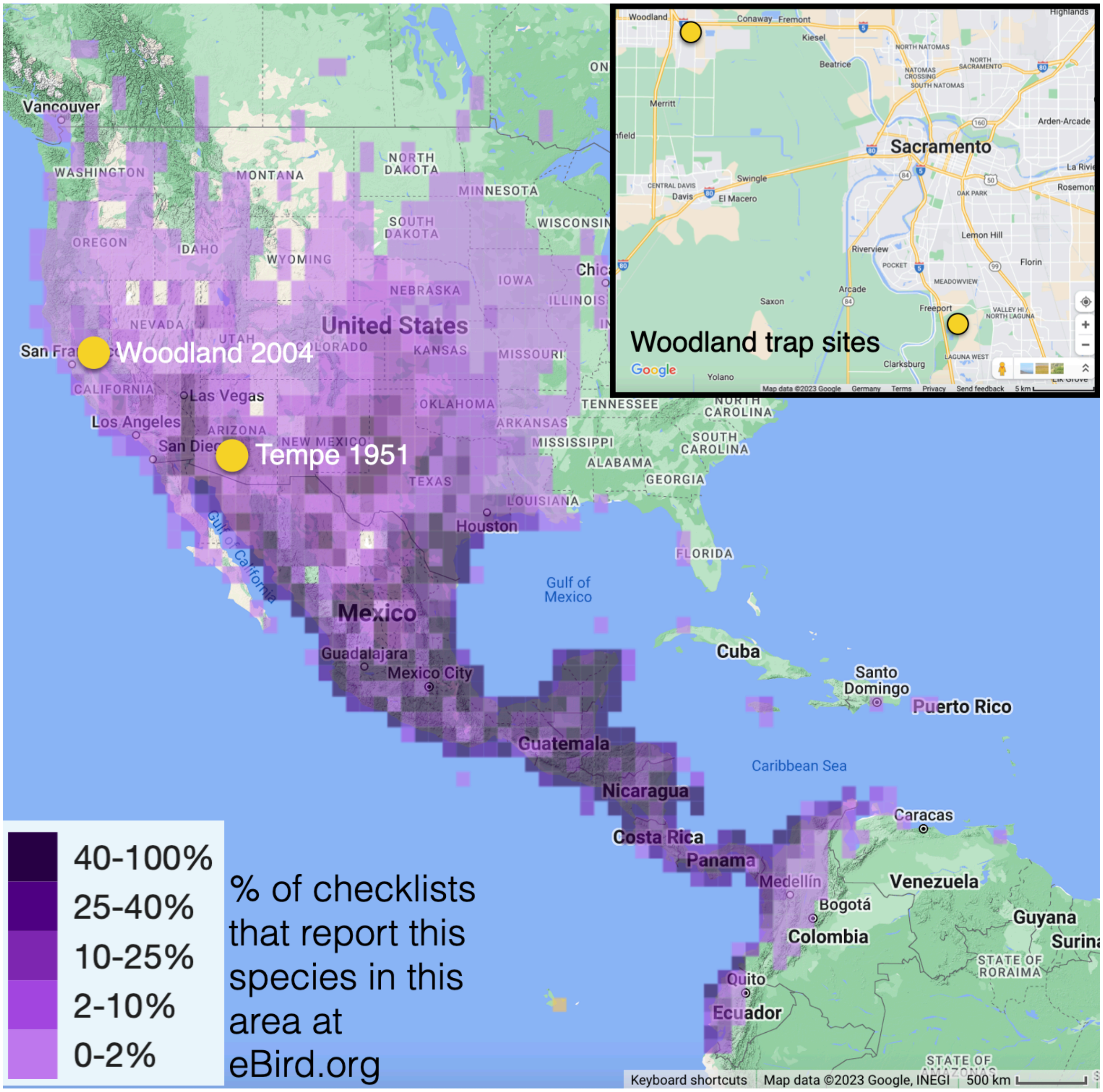

Figure 1: Great-tailed grackle field sites: Woodland is a recently established population (first breeding at the trapping location recorded in 2004) on the northern edge of the range, and Tempe is an older population (established in 1951) in the middle of the northern expansion front. Data from eBird.org.

Table 1. Population characteristics for the field sites. The number of generations at a site is based on a generation length of 5.6 years for this species (BirdLife_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* from July 2004; and Tempe, Arizona: estimated based on 1945 first-sighting report in nearby Phoenix, Arizona (Wehtje, 2004) to which we added 6 years to account for the average time between first-sighting and first-breeding - see Table 3 in Wehtje (2003). The average number of generations was calculated using the number of years of breeding (the "Breeding since" year up to 2020, the final year of data collection in Tempe, and 2022, the final year of data collection in Woodland) divided by the 5.6 year generation length.

Site	Range position	Breeding since	Number of years breeding	Average number of generations	Citation
Tempe, Arizona	Middle of expansion	1951	69	12.3	Wehtje 2003, 2004
Woodland, California	Northern edge	2004	18	3.2	Burrowing Owl July 2004, Pandolfino et al. 2009

RESEARCH QUESTION: Are there differences in behavioral traits (flexibility, innovation, exploration, and persistence) between populations across the great-tailed grackle's geographic range?

Prediction 1: If behavior modifications are needed to adapt to new locations, then there is a higher average and/or larger variance of at least some traits (behaviors) 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 population** (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 will 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 supports the hypothesis* that changes in behavioral traits facilitate the great-tailed grackle's geographic range expansion.

METHODS

Sample

Great-tailed grackles were caught in the wild in Woodland and in the Bufferlands of Sacramento, California. Some of our banded individuals were found in Woodland and the Bufferlands, which are 32 km apart, therefore we considered this one population. We caught grackles with walk-in traps and mist nets. Mist nets decrease the likelihood of a selection bias for exploratory and bold individuals because grackles cannot see the trap. We aimed 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., early-life experience plays a role in the development of holding/grasping objects, Collias & Collias, 1964; Rutz et al., 2016) with variation in our target variables of interest. Observations from members of the Yolo Audubon Society in Woodland, Davis, and Sacramento, California suggest that movement into new areas is most likely by adults or groups of mixed age individuals (Yolo Audubon Society's newsletter *The Burrowing Owl*). Accordingly, if there are differences associated with presence at the edge of the rane, these differences should also be expressed in adults. Adults were identified from their eye color, which changes from brown to yellow upon reaching adulthood (Johnson & Peer, 2001). However, due to difficulties in trapping this species at this site, we also tested some juveniles. This should not pose a problem because we found that the two juveniles (Taco and Chilaquile) we tested in the Tempe population did not perform differently from adults (Blaisdell et al., 2021a; Logan et al., 2021; Seitz, 2021; Logan et al., 2023a). We applied colored leg bands in unique combinations for individual identification. Some individuals (n=33 in Woodland) were brought temporarily into aviaries for behavioral choice tests, and then released back to the wild at their point of capture. Grackles were individually housed in an aviary (each 244 cm long by 122 cm wide by 213 cm tall) for 3 weeks to 6 months where they had *ad lib* access to water at all times and were fed Mazuri Small Bird maintenance diet *ad lib* during non-testing hours (minimum 20 h per day), and various other food items (e.g., peanuts, bread) during testing (up to 4 h per day per bird). Individuals were given three to four days to habituate to the aviaries and then their test battery began on the fourth or fifth day (birds were usually tested six days per week, therefore if their fourth day occurred on a day off, they were tested on the fifth day instead).

We tested as many great-tailed grackles as we could during the 2 years we spent at each of our field sites given that the birds were only brought into the aviaries during the non-breeding season (September through April). It is time intensive to conduct the aviary test battery (3 weeks-6 months per bird), therefore we aimed to meet the minimum sample sizes in Supplementary Material 1 and 2. We aimed for an equal sex ratio of subjects (50% female) and achieved an overall 47% female (this percentage differed depending on the test). We expected to test 20 grackles per site. See the gxpopbehaviorhabitat_data_testhistory.csv data sheet at Logan et al. (2023c) for a list of the order of experiments for each individual at the Woodland site, and g_flexmanip_data_AllGrackleExpOrder.csv at Logan et al. (2023b) for the Tempe grackles.We stopped collecting data on wild-caught great-tailed grackles once we met our minimum sample size (Supplementary Material 1 and 2).

Protocols

Experimental and habituation protocols are available in Supplementary Material 5. In brief, the **flexibility** protocol (from Logan et al., 2023a) used reversal learning with color tubes. Grackles were first habituated to a yellow tube and trained to search for hidden food. A light gray tube and a dark gray tube were placed on the table or floor: one color always contained a food reward (not visible by the bird) while the other color never contained a reward. The bird was allowed to choose one tube per trial. An individual was 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 used a sliding window in 1-trial increments to calculate whether they passed after their first 20 trials. Once a bird learned to prefer one color, the contingency was reversed: food was always in the other color and never in the previously rewarded color. The flexibility measure was how many trials it took to reverse their color preference using the same passing criterion. The first rewarded color in reversal learning was counterbalanced across birds. The rewarded option was pseudorandomized for side (and the option on the left was always placed on the substrate first by the experimenter). Pseudorandomization consisted 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, was 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 were assigned randomly to any given 10-trial session using a random number generator (random.org) to generate a number from 1-88.

The **innovativeness** protocol (from Logan et al., 2023a; and based on the experimental design by Auersperg et al., 2011) used a multiaccess log. Grackles were first habituated to the log apparatus with all of the doors locked open and food inside each locus. After habituation, the log, which had four ways of accessing food (pull drawer, push door, lift door up, swing door out), was placed on the ground and grackles were allowed to attempt to solve or successfully solve one option per trial. Once a bird successfully solved an option three times, it became non-functional (the door was locked open and there was no food at that locus). The experiment ended when all four loci became non-functional, if a bird did not come to the ground within 10 min in three consecutive test sessions, or if a bird did 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 was measured as the proportion of trials participated in during the flexibility and innovativeness experiments (after habituation, thus it is not confounded with boldness). The higher the number, the more persistent they were. This measure indicates that those birds who do not participate as often were less persistent in engaging with the task. We generally offered a grackle the chance to participate in a trial for 5 min. If they did not participate within that time, we recorded -1 in the data sheet, the apparatus was removed, and the trial was re-attempted later.

Exploration was measured as the latency to approach within 20 cm of a novel environment inside of their familiar aviary environment and this test was conducted two times for each bird so we could obtain individual consistency measures. Time 1 occurred on the individual's 8th day in the aviary and Time 2 occurred 1 week after Time 1. The bird's regular food was moved to one end of the aviary, away from the novel environment, and we first conducted a motivation test where we placed a piece of preferred food on the ground and waited out of view for 5 min. We only proceeded with the exploration assay if the bird ate the food. This motivation test allowed us to determine whether the grackle was interested in coming to the ground at all, where, for example, a grackle might not eat the food because it has just bathed and is primarily focused on preening and drying feathers. The bird was then exposed to first a familiar environment without the novel environment for 45 min and then to a novel environment (a tent) that is placed on the ground within the familiar environment for 45 min. If an individual did not approach within 20 cm, it was given a latency of 2701 sec (45 min plus 1 sec). In a previous experiment (McCune et al., 2019b), we validated that grackles did not perceive the novel environment as threatening (i.e., it was not a measure of boldness).

Experimental order: The order of experiments for reversal learning or multiaccess log, was counterbalanced across birds for the Woodland population. The Arizona population received the reversal learning experiment first because their flexibility was manipulated to determine whether manipulating flexibility influences performance on subsequent tests (see Logan et al., 2023a).

Statistical analyses

We used **simulations** and designed **customized models** to determine what sample sizes would allow us to detect differences between sites following the methods in McElreath (2020a, Supplementary Material 1 and 2; see chapter 5.3 in Bolker, 2008 for why simulations perform more powerful power analyses). We did not **exclude** any data, and if data were **missing** (e.g., if a bird participated in only one of the two experiments) for an individual in a given experiment, then this individual was not included in that analysis. Analyses were conducted in R (current version ‘r getRversion(), R Core Team, 2023) and Stan (version 2.18, Carpenter et al., 2017) using the following packages: psych (Revelle, 2017) and irr (Gamer et al., 2012) for calculating interobserver reliability scores; rethinking (McElreath, 2020a), cmdstanr (Gabry & Češnovar, 2021), rstan (Stan Development Team, 2020), posterior (Vehtari et al., 2021) and Rcpp (Eddelbuettel & François, 2011) for conducting Bayesian analyses; knitr (Xie, 2013, 2017, 2018), formatR (Xie, 2023), dplyr (Wickham et al., 2021), tidyr (Wickham et al., 2023), kableExtra (Zhu, 2021), lattice (Sarkar, 2008), and gridExtra (Auguie, 2017) for formatting; DHARMa (Hartig, 2019) for data cleaning; lme4 (Bates et al., 2012; Bates et al., 2015) and MCMCglmm (Hadfield, 2010a) for running GLMMs; and rptR (Stoffel et al., 2017) for calculating repeatability. Interobserver reliability scores indicated high agreement across coders for all dependent variables (see Supplementary Material 3 for details).

Flexibility analyses

Model and simulation

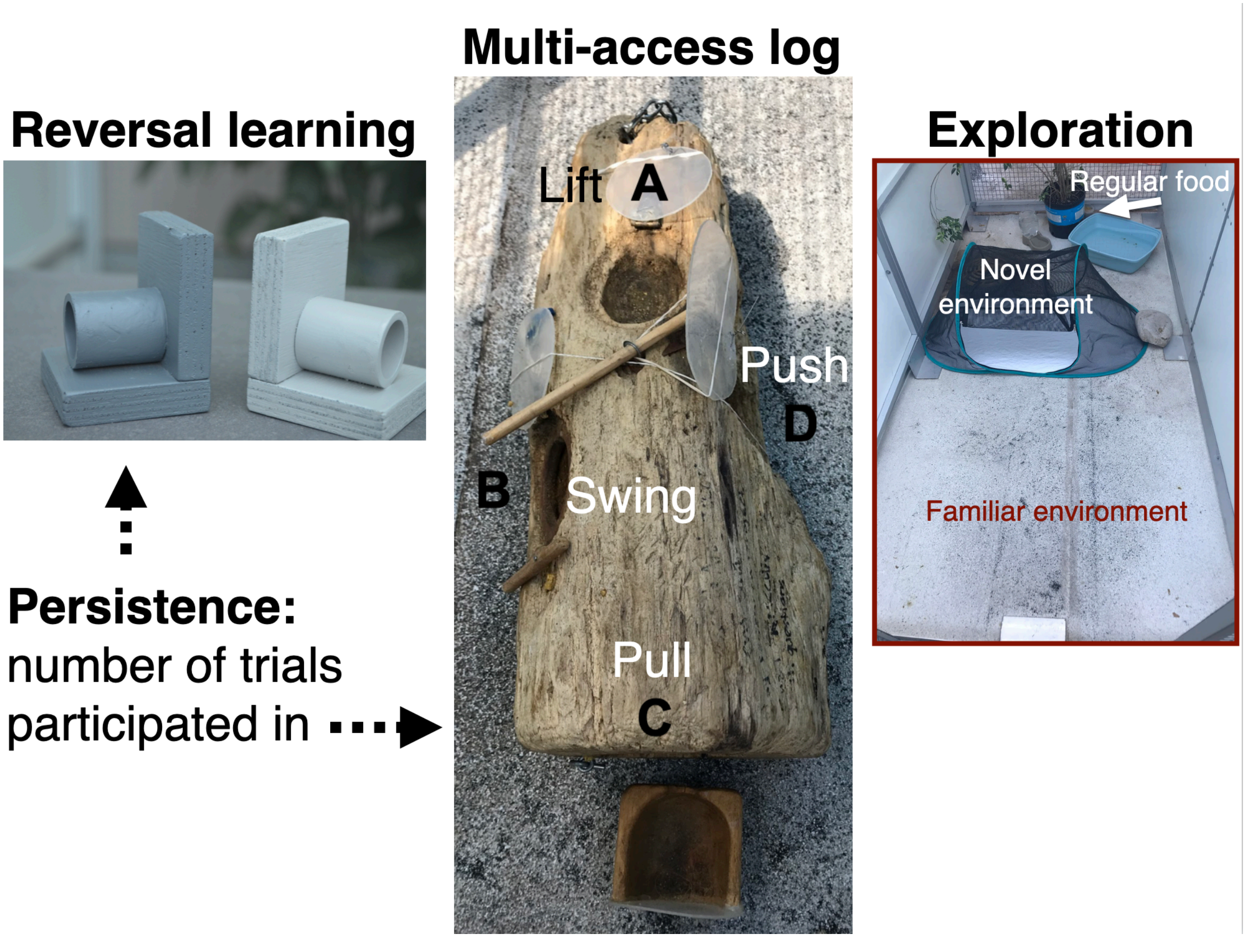

Figure 2: Experimental protocol (see Supplementary Material 5 for more details). Great-tailed grackles from the older and newer populations were tested for their: (top left) flexibility as the number of trials to reverse a previously learned color tube-food association; (middle) innovativeness as the number of loci (lift, swing, pull, push) solved to obtain food from within a multiaccess log; (bottom left) persistence as tge proportion of trials participated in during flexibility and innovativeness tests; and (far right) exploration as the latency to approach a novel environment placed inside of the familiar environment with regular food present, but not near the novel environment. The order of the flexibility and innovativeness experiments was counterbalanced for the California grackles and they received their first exploration assay as close as possible to day 8 in the aviaries. The Arizona grackles received the flexibility experiment first (because they underwent a flexibility manipulation) and the innovativeness experiment and exploration assay afterward (note that there could have been other experiments between the flexibility experiment and the innovation experiment and exploration assay because their test battery was much larger than that of the California birds, Logan et al., 2023a). See the test history for each bird in the gxpopbehaviorhabitatq1_data_testhistory.csv data sheet at Logan et al. (2023c).

We modified the reversal learning Bayesian model in Blaisdell et al. (2021a) to simulate and analyze population differences in reversal learning, and calculated 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 is made depending on whether that choice contains a food reward or not. It does this by updating three main components for each choice: an attraction score (how much an individual prefers one option over the other), a learning rate (ϕ; higher values mean the individual updates their attraction score at a higher rate), and a rate of deviating from learned attractions (λ; lower values mean the individual is choosing between the options more randomly). The attraction score is the weight an individual gives to a particular option based on its past reward history for that option with attractions increasing if they received a reward when previously choosing that option. The decision regarding which of the two options to make is determined by the relative weights of the two attraction scores (each option gets its own attraction score).

As in Blaisdell et al. (2021a), we, too, used previously published data on reversal learning of color tube preferences in great-tailed grackles in Santa Barbara, California (Logan, 2016a) to inform the model modifications. We modified the Blaisdell et al. (2021a) model in a two 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 (see Lukas et al., 2022 for more detail). 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 are 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 only changes 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; see Lukas et al. (2022) for more detail]. Previously, both attractions were updated after every trial, assuming that individuals understand that the experiment is set up 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 ϕ to have the same learning rate as a model where both attraction scores update after every trial (Lukas et al., 2022). This change also appears to better reflect the performance of the Santa Barbara grackles, because they had higher ϕ values, which, in turn, meant lower λ values to reflect the performance during their initial learning. These lower λ values better reflected the birds' behavior during the first reversal trials: a large λ value means that birds continue to choose the now unrewarded option almost 100% of the time, whereas the lower λ values mean that birds start to explore the rewarded option relatively soon after the switch of the rewarded option (Lukas et al., 2022).

We first reanalyzed the Santa Barbara grackle data to obtain the ϕ and λ 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 model. This model was used to analyze the actual data after it was collected, using only data from the first reversals to eliminate the need to modify the model to include treatment (whether an Arizona grackle was manipulated or not). We used an analysis called a contrast to assess whether one site was systematically larger or smaller than the other by estimating what percentage of each sample of differences is either larger or smaller than zero. If 89% of the differences are larger than zero, then the older population has a larger mean, and if 89% of the differences are smaller than zero, then the edge population has a larger mean. If 89% of the differences cross zero, then we conclude that there is no strong difference between the sites. See Supplementary Material 1 and 2 for more details. To determine whether there were differences between the variances in ϕ and λ between sites, we conducted models as follows:

ϕ_i or λ_i ~ Normal(μ, σ[site]) [likelihood],

log(μ) ~ α[site] [model],

where either ϕ_i or λ_i were used as the response variable, σ[site] allows a separate variance to be assigned to each site, α is the intercept for the ϕ_i or λ_i means, and each site gets its own intercept. We then ran a contrast to determine whether there was a difference in variances between the sites.

Innovation analysis

Model and simulation

Expected values for the number of loci 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, *Aphelocoma californica*, and Mexican jays, *Aphelocoma wollweberi*: McCune, 2018; McCune et al., 2019a).

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

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

where $locisolved_i$ 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, α is the intercept, and each site gets its own intercept, and β is the slope between the probability of solving a locus and the *treatment* (flexibility manipulated or not). After running simulations, we identified the following distribution to be the most likely priors for our expected data:

$\alpha \sim \text{Normal}(\bar{a},\sigma)$ *[α prior]*

We used a normal distribution for α because it is a sum (see Figure 10.6 in McElreath, 2020a) and a logit link to ensure the values are between 0 and 1. We set the mean to $\bar{a}$ and the standard deviation to σ to allow the model to learn from the first site it analyzes and apply that learning to the next site (called partial pooling, McElreath, 2020a). We again used a contrast analysis (McElreath, 2020a) to assess whether one site was systematically larger or smaller than the other by estimating what percentage of each sample of differences is either larger or smaller than zero. See Supplementary Material 1 and 2 for more details.

We modified the above model to analyze the variance in loci solved between sites by adding c[individual],

$\text{logit}(p) \sim \alpha[\text{site}] + \beta * treatment + \text{c[individual]}$ *[model]*,

which gives the proportion of loci solved per bird. We specified the priors for this as

$\text{c[individual]} \sim \text{dnorm}(0,\sigma[\text{site}])$,

$\sigma[\text{site}] \sim \text{dexp}(1)$,

where $\sigma[\text{site}]$ gives the average variance per site. We then conducted a contrast analysis to determine whether sites differed.

Note that two grackles, Kau and Galandra, were accidentally able to pull 2 and 1, respectively, locus doors open during habituation to the multiaccess box. Because habituation was not observed by an experimenter, the birds had the possibility to learn how these doors worked. Therefore, these doors were locked open and non-functional throughout their entire experiment. We accounted for this in the model by replacing the 4 (as in 4 possible loci were available to solve) with a column of data that listed the maximum possible loci available to each bird.

Exploration analysis

Model and simulation

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

$latency_i \sim \text{gamma-Poisson}(L_i, P)$ *[likelihood]*,

$\log(L_i) \sim \alpha[\text{site}] + \beta * treatment$ *[the model]*,

where $latency_i$ is the average latency to approach a novel environment, L_i is the rate (probability of approaching the novel environment in each second) per bird (and we took the log of it to make sure it was always positive; birds with a higher rate have a smaller latency), P is the dispersion of the rates across birds, α is the intercept for the rate per site, and β is the slope between the $latency_i$ and the *treatment* (flexibility manipulated or not).

Expected values for the latency to approach a novel environment range from 0-2700 sec, which encompassed the time period during which they were exposed to the novel environment (sessions lasted up to 45 min). However, we did 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 novel environment 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:

P ~ 1/(Exponential(1)) *[P prior]*,

α[site] ~ Normal($\bar{a}$,σ) *[α prior]*.

We used a gamma-Poisson distribution for latency because it constrains the values to be positive. For P, we used an exponential distribution because it is standard for this parameter. We used a normal distribution for α[site] because it is a sum with a large mean (see Figure 10.6 in McElreath, 2020b), and we set the mean to $\bar{a}$ and the standard deviation to σ to allow the model to learn from the first site it analyzes and apply that learning to the next site (called partial pooling, McElreath, 2020a). We used a contrast analysis (McElreath, 2020a) to assess whether one site was systematically larger or smaller than the other by estimating what percentage of each sample of differences is either larger or smaller than zero. See Supplementary Material 1 and 2 for more details.

To analyze variance in exploration between sites, we conducted a right-censored model because it was better able to manage the many cases in the Woodland population where birds never approached the novel environment and therefore had latency values of 2701 sec (McElreath, 2020a). The model is as follows:

$latency_i$ | latency<=2700 ~ exponential(L),

which indicates that the bird approached the novel environment (the event happened),

$latency_i$ | latency==2701 ~ custom(exponential_lccdf(!Y|L)),

which indicates that the bird did not approach (the event did not happen),

L <- 1.0/μ,

log(μ) <- α[site] + β * *treatment* + c[individual],

c[individual] ~ dnorm(0,σ[site]),

where μ is the average $latency_i$, L is the log average time to approach novel environment, L gets a different rate for each site (α[site]) and for each individual (c[individual]), and β is the slope between L and the *treatment* (flexibility manipulated or not). The offsets for each individual, c[individual], from the site mean (α[site]), are also clustered by site, σ[site], to determine the variance among individuals at each site. We then ran a contrast to determine whether there was a difference in variances between the sites.

Persistence analysis

Model and simulation

Expected values for the number of trials not participated in could range from 0-125. The likely maxima for reversal learning is 300 trials based on data from Santa Barbara (Logan, 2016b) and Tempe grackles (Logan et al., 2023a). On average, individuals participated in 70 trials in the initial discrimination, a maximum of 130 trials in the reversal, and up to 100 non-participation trials across the initial discrimination and reversal. On the multiaccess log, grackles participated in a maximum of 50 trials and there were up to 25 non-participation trials. The estimated maximum number of non-participation trials is based on what might be expected from an individual who does not participate very often. After running simulations, we identified the following distribution and priors as most likely for our expected data:

$participated_i$ ~ Binomial($totaltrials_i$, p) *[likelihood]*,

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

where $participated_i$ indicates whether the bird participated or not in a given trial, $totaltrials_i$ 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, α is the intercept, and each site gets its own intercept, and β is the slope between whether the individual participated or not ($participated_i$) and the *treatment* (flexibility manipulated or not). We used a logit link to constrain the output to between 0 and 1. After running simulations, we identified the following distribution and priors as most likely for our expected data:

$\alpha[\text{site}] \sim \text{Normal}(\bar{a},\sigma)$ *[α prior]*.

We used a normal distribution for α because it is a sum (see Figure 10.6 in McElreath, 2020a). We set the mean to $\bar{a}$ and the standard deviation to σ to allow the model to learn from the first site it analyzes and apply that learning to the next site (called partial pooling, McElreath, 2020a). We used a contrast analysis [rethinking2020] to assess whether one site was systematically larger or smaller than the other by estimating what percentage of each sample of differences is either larger or smaller than zero. See Supplementary Material 1 and 2 for more details. See the Innovation analysis section for how we analyzed the variance in the proportion of trials participated in - it is the same model but replaces loci solved with proportion of trials participated in.

Repeatability of exploration and persistence

We obtained repeatability estimates that account for the observed and latent scales, and then compared 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 ran this GLMM using the MCMCglmm function in the MCMCglmm package (Hadfield, 2010b) with a Poisson distribution and log link using 13,000 iterations with a thinning interval of 10, a burnin of 3,000, and minimal priors (V=1, nu=0) (Hadfield, 2014). We ensured the GLMM showed acceptable convergence (i.e., lag time autocorrelation values <0.01, Hadfield, 2010b), and adjusted parameters if necessary.

Post-study choices made since receiving in principle recommendation

This study began as a preregistration that was peer reviewed and received in principle recommendation at PCI Ecology in 2019 (Logan et al., 2019). While our ideal plan was to conduct the same tests at an additional field site in Central America, due to restrictions around COVID-19 and also to issues with sexual abuse at the planned field site, it was not possible for us to accomplish this goal within our current funding period.

In the preregistration, we said that for the exploration measure we would use the "Latency to approach within 20 cm 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)." We had data for both exploration measures and we used the latency measure because this was the variable that our preregistered analysis was designed for.

In the peer review history of the preregistration, we said that we would use whichever exploration test was repeatable with the Tempe grackles (novel object and/or novel environment) (round 1, response 16, https://ecology.peercommunityin.org/articles/rec?id=98). The methods for both novel stimuli were exactly the same and there was little variation in whether, or for how long, individuals went into the novel environment (i.e., most individuals did not go in the novel environment). However, the Tempe grackles responded differently to the novel environment and novel object, therefore they did not perceive the stimuli as the same. From the Tempe grackle data, we found that responses were only repeatable for the novel environment test (McCune et al., 2019b). Therefore, we conducted this assay (and not the novel object assay) with the Woodland grackles and compared the two populations on this one assay.

For the repeatability of persistence, the preregistered model had Test (reversal or multiaccess box) as the explanatory variable and ID as the random variable. However, we believe we made an error in choosing the explanatory variable because we are interested in whether the trait is repeatable across populations regardless of the test. Therefore, we replaced Test with Population in the model. In addition, we realized that our measure of persistence (proportion of trials participated in) is not appropriate for a Poisson model, as preregistered. Consequently, we used a likelihood ratio test to compare a mixed model to a model without the ID random effect, and the function rpt from the package: rptR (Stoffel et al., 2017) to estimate the variance in the dependent variable attributable to consistent differences among individuals across the two tests. We previously found that this method produces the same repeatability results as the MCMCglmm method using a Gaussian distribution (McCune et al., 2022).

The exploration data for the repeatability calculation were heteroscedastic and overdispersed. Additionally, 53% of the data were at the ceiling value (i.e., the bird did not approach the novel environment). Consequently, the model that best fit the data and was appropriate for the repeatability analysis was a binomial model, where the response was 0 (the grackle never approached the novel environment during exploration trials) or 1 (the grackle approached the novel environment).

RESULTS

See Table 2 for summary results for grackles in Woodland and Tempe, as well as some data for boat-tailed grackles (population: BTGR), which we describe in the Discussion.

Table 2. Summary data by bird for each of the variables measured. Population indicates where they were trapped (Sacramento is part of the Woodland population), bird is the bird's name, sex indicates whether they are female (F) or male (M), learn speed is the number of trials to form the initial color preference, reversal speed is the number of trials to reverse the color preference (first reversal), ϕ and λ are the two flexibility components (from the first reversal), MAB loci are the number of loci solved on the multiaccess box, MAB max is the maximum number of loci available to that bird, Exploration is the average number of seconds it took the bird to approach within 20 cm of the novel environment (note that 2701 s means the bird never approached), persistence is the proportion of reversal learning and multiaccess box trials the bird participated in, and flexibility manipulated indicates whether this was one of the 8 Tempe grackles who underwent the serial reversals to make them more flexible (Yes) or not (No). "X" indicates that this bird did not complete this experiment or that we cannot count the data for this experiment, and "-" indicates this bird was not given this experiment.

Population	Bird	Sex	Learn speed	Reverse speed	Phi	Lambda	MAB loci	MAB max	Exploration	Persistence	Flexibility manipulated
Sacramento	Flan	F	37	67	0.0808	4.37	4	4	1460.2	1.00	No
Sacramento	Tembleque	M	27	39	0.0496	6.12	4	4	636.1	0.57	No
Sacramento	Polvorones	M	23	47	0.0693	2.75	3	4	802.5	0.63	No
Sacramento	Alegria	F	46	60	0.0216	6.66	1	4	2701	0.48	No
Sacramento	Piña	F	46	77	0.0283	6.73	4	4	2701	0.97	No
Sacramento	Camote	M	31	67	0.0252	3.76	4	4	2701	0.88	No
Sacramento	Helado	M	29	82	0.0163	7.79	4	4	2701	0.77	No
Sacramento	Zapote Negro	M	30	70	0.0238	4.47	2	4	2701	0.90	No
Sacramento	Buñuelo	M	47	92	0.0334	3.38	4	4	335.2	0.99	No
Sacramento	Dulce de Leche	M	32	66	0.0331	4.85	4	4	2701	0.98	No
Sacramento	Kau	M	38	79	0.0483	2.85	2	2	308.6	0.89	No
Woodland	Galandra	JF	50	96	0.0573	4.06	3	3	2005.1	0.88	No
Woodland	Kel	F	61	64	0.0133	5.06	3	4	2701	0.75	No
Woodland	Ak'xi'	M	47	76	0.0160	5.17	3	4	1959.3	0.80	No
Sacramento	Cuervo	JM	31	49	0.0122	7.31	4	4	722.8	0.85	No
Sacramento	Xunub	M	76	100	0.1761	2.99	4	4	2701	0.78	No
Sacramento	Cocinera	F	35	82	0.0049	8.18	1	4	2701	0.90	No
Sacramento	Tzanatl preciosa	F	21	26	0.0217	5.09	4	4	2701	0.59	No
Sacramento	Cutuy	F	40	150	0.0431	5.44	3	4	1088.1	0.97	No
Sacramento	Xango	M	X	-	-	-	0	4	1235.3	0.28	No
Sacramento	Wachil	JF	X	-	-	-	1	4	2701	0.33	No
Sacramento	Talingo	M	X	-	-	-	4	4	2027.7	0.73	No
Sacramento	Quiscalus	F	X	-	-	-	3	4	471.3	0.46	No
Sacramento	Churro	M	-	-	-	-	-	-	1951.8	-	No
Sacramento	Chocolate	F	X	-	-	-	-	-	2701	0.34	No
Sacramento	Sopapilla	F	-	-	-	-	-	-	1591	-	No
Sacramento	Tres Leches	F	-	-	-	-	-	-	2196.8	-	No
Sacramento	Merengue	M	-	-	-	-	-	-	2701	-	No
Sacramento	Carlota	F	X	-	-	-	-	-	1786.9	0.32	No
Sacramento	Changa	F	-	-	-	-	-	-	549.0	-	No
Sacramento	Urraca	M	-	-	-	-	-	-	2701	-	No
Sacramento	Bacmut bacni	M	X	-	-	-	-	-	1530.6	-	No
Tempe	Tomatillo	M	35	50	0.0154	5.04	-	-	431	0.64	No
Tempe	Queso	M	48	68	0.0257	4.88	-	-	1943.4	0.87	No
Tempe	Tapa	F	28	98	0.0572	3.03	-	-	1616.9	0.80	No
Tempe	Yuca	F	31	80	0.0456	3.22	4	4	1816.3	0.98	No
Tempe	Marisco	M	34	48	0.0727	4.59	2	4	2701	0.88	No
Tempe	Pizza	M	45	59	0.1018	5.55	1	4	2701	0.75	No
Tempe	Mofongo	M	20	38	0.0530	4.20	4	4	71.8	0.47	No
Tempe	Taquito	M	81	159	0.0380	4.09	4	4	2701	0.94	No
Tempe	Chalupa	F	49	90	0.0362	2.58	-	4	1848.4	0.72	Yes
Tempe	Mole	M	22	70	0.0139	5.57	4	4	73.5	0.95	Yes
Tempe	Habanero	M	41	78	0.0213	5.39	-	-	351.6	0.83	Yes
Tempe	Diablo	M	20	80	0.0097	5.52	1	4	2701	0.82	Yes
Tempe	Burrito	M	29	59	0.0564	3.64	4	4	70.4	0.89	Yes
Tempe	Adobo	M	50	100	0.0264	3.68	4	4	446.9	1.00	Yes
Tempe	Chilaquile	JM	22	39	0.0192	5.95	4	4	1422.7	1.00	Yes
Tempe	Pollito	M	35	58	0.0084	5.56	3	4	909.8	0.57	Yes
Tempe	Taco	JM	37	78	0.0528	2.97	4	4	148.3	0.89	No
Tempe	Memela	F	38	59	0.0223	4.37	-	-	-	0.55	No
Tempe	Fideo	M	60	70	0.0206	5.19	-	-	2701	0.66	No
Tempe	Avocada	F	50	99	0.0159	8.77	-	-	677.7	0.72	No
BTGR	Verbena	M	20	35	0.0108	4.35	0	4	2701	0.72	No
BTGR	Pawpaw	M	20	79	0.0682	6.00	0	4	2701	0.75	No
BTGR	Bladderwort	M	82	X	-	-	4	4	2701	0.71	No
BTGR	Roseling	M	X	-	-	-	1	4	2701	0.36	No
BTGR	Cocoplum	M	27	-	-	-	-	-	2701	0.71	No

Flexibility

There were no strong site differences for either the ϕ or λ component of reversal learning (using data from the first reversal): ϕ or λ (Figure 3). However, the average ϕ per population differed by only 0.0012 (Woodland=0.0313, Tempe=0.0301) and λ by 0.29 (Woodland=4.80, Tempe=4.51), and the compatibility intervals for the estimated differences for both parameters in the contrast analysis crossed zero (Table 3; n=19 birds in Woodland, n=19 birds in Tempe). With our sample size, we only have the power to reliably detect differences between the populations if they are larger than 0.01 for ϕ, which corresponds to a difference of 1% in how much individuals choose the rewarded option after they have just received a reward from this option. For λ, we would need a difference of at least 3, which corresponds to a 10% difference in how often an individual chooses the alternative option.The detection differences in ϕ and λ are based on our power analysis in Supplementary Material 2, summarized in Supplementary Material 1, and their correspondence with the number of trials to reverse comes from Blaisdell et al. (2021b). Accordingly, we cannot exclude that the two populations are different, however we can estimate the range for how small the difference can be. Based on the estimated 89% compatibility intervals (McElreath, 2020a) for ϕ and λ in Table 3, the two populations are unlikely to differ by more than 0.01 for ϕ and 1.48 for λ. Woodland grackles had a larger variance in ϕ (mean=0.02, standard deviation=0.01, 89% compatibility interval=0.0009-0.03) than the Tempe grackles, and there were no strong differences in variance in λ (mean=0.26, sd=0.39, 89% CI=-0.37-0.88), as indicated by the contrast analyses.

Table 3. Contrasts (indicated by "diff") between populations for the flexibility measure of reversal learning: ϕ and λ (data from the first reversal).

	Mean	Standard deviation	Lower 89 percentile compatibility interval (5.5%)	Upper 89 percentile compatibility interval (94.5%)
Woodland Phi	0.04	0.04	0.01	0.08
Tempe Phi	0.04	0.02	0.01	0.07
diff_Phi	0.00	0.01	-0.01	0.01
Woodland Lambda	5.11	1.66	2.84	7.80
Tempe Lambda	4.69	1.40	2.98	5.94
diff_Lambda	0.29	0.68	-0.75	1.48

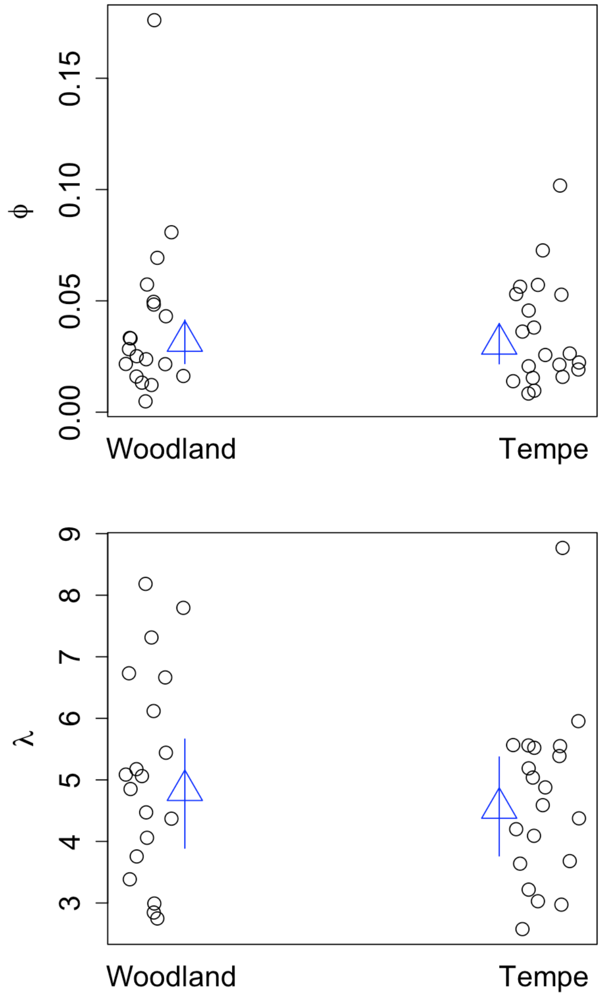

Figure 3. Measures of flexibility from the first reversal of the reversal learning experiment: ϕ and λ per individual in each population. The black circles are the raw data from each bird, the blue triangles are the population means, and the blue lines are their 89% compatibility intervals.

Innovation

There were no differences in innovativeness between the sites: individuals at both sites solved similar proportions of loci on the multiaccess box as indicated by the contrast that showed that the compatibility interval crossed zero (diff_12 in Table 4; Figure 4; Woodland: n=23 birds, mean loci solved=3.0; Tempe: n=12 birds, mean loci solved=3.25). We would need a difference of at least 0.8 to 1.0 loci solved to detect a difference between the sites (based on our power analysis in Supplementary Material 2, summarized in Supplementary Material 1). However, the number differed by only 0.25 (Table 4). We found no support that the variances differ between the two populations because the contrast analysis showed the compatibility interval crossed zero (mean=-0.07, sd=1.08, 89% CI=-1.89-1.50).

Table 4. Contrasts between populations for the innovation measure: the proportion of loci solved on the multi-access box.

	Mean	Standard deviation	Lower 89 percentile compatibility interval (5.5%)	Upper 89 percentile compatibility interval (94.5%)
Woodland	0.78	0.04	0.71	0.83
Tempe	0.78	0.06	0.68	0.87
diff_12	-0.01	0.06	-0.11	0.09

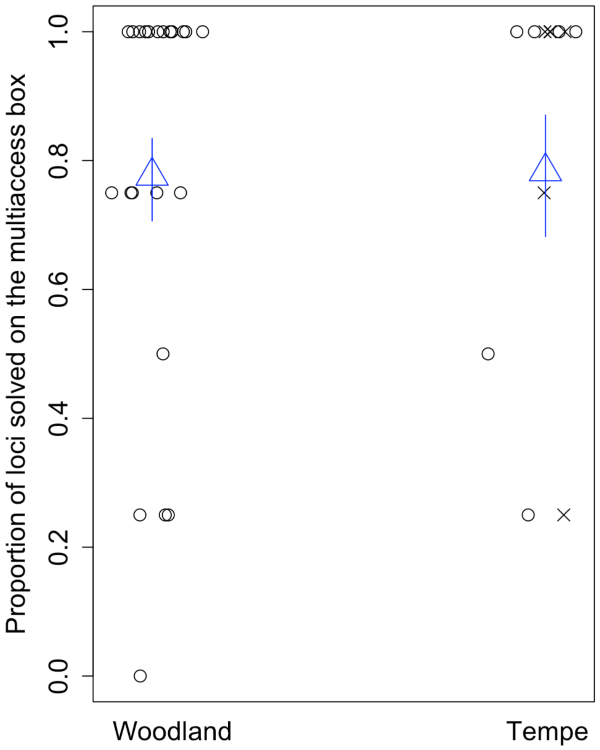

Figure 4. Proportion of loci solved on the multiaccess box in the innovativeness test per individual at each site (n=23 birds in Woodland, n=12 birds in Tempe). The black circles are the raw data from the non-flexibility manipulated birds, the black X's are the flexibility manipulated birds, the blue triangles are the population means, and the blue lines are their 89% compatibility intervals.

Exploration

There were no strong site differences for exploration, which was quantified as the latency to approach within 20 cm of a novel environment (averaged across Time 1 and Time 2; Woodland: n=32 grackles, mean latency=1900 sec, standard deviation=270; Tempe: n=19 grackles and 8 of these were in the flexibility manipulation, mean latency=1641 sec, standard deviation=427) as indicated by the contrast that shows that the compatibility interval crosses zero (diff_12 in Table 5; Figure 5). We would need a difference of more than 824 sec in the latencies to detect a difference between the sites (based on our power analysis in Supplementary Material 2, summarized in Supplementary Material 1). However, the latencies differ by only 259 sec (Table 5). The mean latencies we found were much higher than those used in the power analyses, which makes it more difficult to detect differences with our data because the averages approach the ceiling of 2700 sec and therefore we lose information on the several birds that timed out (had latencies of 2701 sec).

The variances were similar across sites as indicated by the contrast analysis, which showed the compatibility interval crossed zero (mean=-0.57, sd=0.65, 89% CI=-1.70-0.42).

Table 5. Contrasts (indicated by "diff") between populations for the exploration measure: latency (sec) to approach within 20 cm of a novel environment.

	Mean	Standard deviation	Lower 89 percentile compatibility interval (5.5%)	Upper 89 percentile compatibility interval (94.5%)
Woodland	1899.99	269.63	1535.67	2381.05
Tempe	1640.59	427.03	1102.87	2421.47
P	1.64	0.32	1.20	2.19
diff_12	259.40	495.60	-575.89	974.61

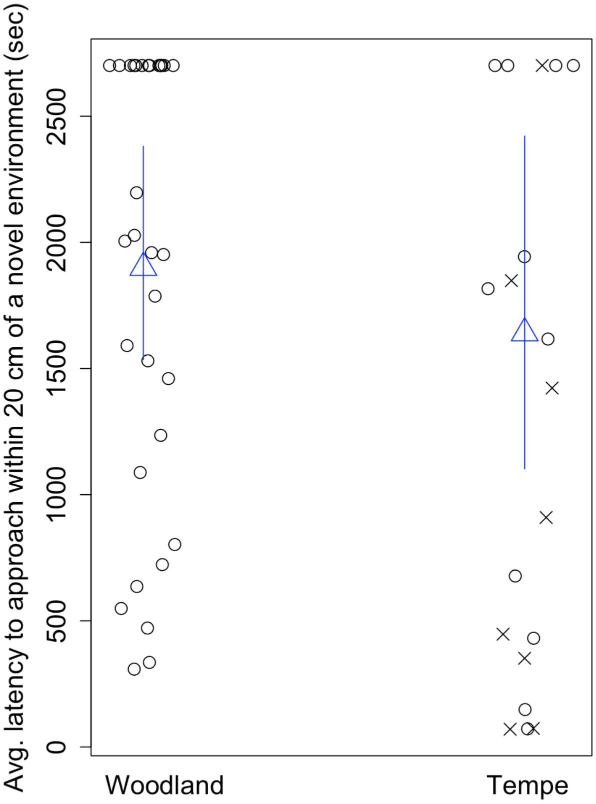

Figure 5. Average latency to approach within 20 cm of a novel environment in the exploration assay per individual at each site (n=32 Woodland, n=19 Tempe and 8 of these were flexibility manipulated). Note that if an individual does not approach within 20 cm of the novel environment at Time 1 or 2, they are given a ceiling value of 2701, which is one second longer than the session length. The black circles are the raw data from the non-flexibility manipulated birds, the black X's are the flexibility manipulated birds, the blue triangles are the population means, and the blue lines are their 89% compatibility intervals.

Persistence

Individuals in the more recent population in Woodland, California were more persistent than those in the older population in Tempe, Arizona (Figure 6; Woodland: n=25 birds, mean proportion of trials participated in=0.78; Tempe: n=20 birds and 8 of these were flexibility manipulated, mean proportion of trials participated in=0.76)). Woodland grackles participated in more of the offered trials in the reversal learning and multiaccess box experiments as indicated by the contrast that shows that the compatibility interval does not cross zero (diff_12 in Table 6). We would need a difference of more than 0.1 in the proportion of trials participated in to detect a difference between the sites (based on our power analysis in Supplementary Material 2, summarized in Supplementary Material 1). The difference we found is less than this at 0.02, which means that this could be a false positive. However, we conducted an analysis to investigate the likelihood of having a false positive and found that it is twice as likely that this is a true positive rather than a false positive (63%; see analysis code in r code chunk "modelpersistence" at the Rmd file). Visual interpretation, through plotting the values (Figure 6), could suggest that the variance in persistence might be larger among the individuals in Woodland compared to Tempe because some of the Woodland individuals show lower persistence values than those in the Tempe individuals. We found no support that the variances differ between the two populations because the contrast analysis showed the compatibility interval crossed zero (mean=0.21, sd=0.40, 89% CI=-0.44-0.83.

Table 6. Contrasts (indicated by "diff") between populations for the persistence measure: proportion of trials participated in across the reversal and multiaccess box experiments.

	Mean	Standard deviation	Lower 89 percentile compatibility interval (5.5%)	Upper 89 percentile compatibility interval (94.5%)
Woodland	0.78	0.01	0.77	0.79
Tempe	0.76	0.01	0.74	0.77
diff_12	0.02	0.01	0.01	0.04

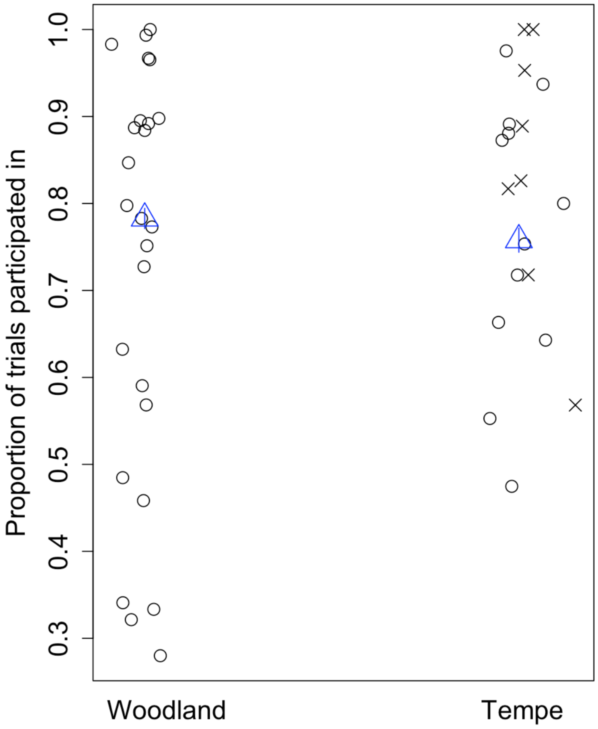

Figure 6. The proportion of trials participated in across the reversal and multiaccess box experiments is the measure of persistence per individual at each site (n=25 Woodland, n=20 Tempe with 8 of these being flexibility manipulated). The black circles are the raw data from the non-flexibility manipulated birds, the black X's are the flexibility manipulated birds, the blue triangles are the population means, and the blue lines are their 89% compatibility intervals.

Repeatability of exploration and persistence

Exploration of the novel environment was repeatable in the Woodland population (current study repeatability (R)=0.70, likelihood ratio test p-value=0.001, confidence interval=0.2-1.0). Our previous analysis found that novel environment exploration was repeatable in the Tempe grackles (McCune et al., 2019b: R=0.72, p<0.001, confidence interval=0.42-0.88). Persistence was repeatable across both populations (R=0.24, p-value=0.03, confidence interval=0.03-0.46).

DISCUSSION

We conducted behavioral experiments with great-tailed grackles from two populations: an older population in the middle of the expansion front in Tempe, Arizona, and a more recent population on the northern edge of their expansion in Woodland, California. We found that individuals in the edge population were more persistent than the population in the middle of the expansion front, and that there are no population differences in behavioral flexibility, innovation or exploration. This supports the hypothesis that changes in particular behaviors are potentially important for facilitating a species' rapid geographic range expansion (Griffin et al., 2017; Szabo et al., 2020). Our measures of flexibility (using serial reversals in the Tempe population, McCune et al., 2022), exploration (Tempe: McCune et al., 2019b, Woodland: reported here), and persistence (both populations reported here) were repeatable and show large inter-individual variation, which validates that these are stable traits that can be meaningfully compared.

We found no support for the hypothesis that a higher average flexibility (reversal learning of a color preference) is required in an edge population (e.g., Lefebvre et al., 1997; Sol & Lefebvre, 2000; Sol et al., 2002, 2005, 2007; Wright et al., 2010; Griffin & Guez, 2014; Chow et al., 2016). That flexibility, the ability to change behavior in reaction to changing circumstances through packaging information and making it available to other cognitive processes, was not on average higher among individuals at the edge of the expansion range indicates that flexibility is not a latent trait that is called upon when individuals move into new areas (Wright et al., 2010). We found that the edge population had a higher variance in one of the two components of flexibility, ϕ, the learning rate. This indicates that there is a larger diversity of this flexibility component in the population, which means that there is a higher chance that at least some individuals in the population could be more flexible (this seems to be driven by a single individual having a particularly high ϕ, see Figure 3). We were unable to find comparable studies of flexibility averages and variances across the range of species that are rapidly expanding their range in which to contextualize our results. However, invasion ecology theory supports the idea that large variance in behavioral traits within species facilitates range expansion or invasion success at multiple points in the invasion process (Chapple et al., 2012). Further experimental research in more species is required to be able to generalize about whether higher flexibility variances are consistently associated with rapid range expansions.

It is possible that behavioral flexibility facilitated the increase of this species' habitat breadth beyond marshes when humans started to modify the environment in central America thousands of years ago (Christensen, 2000). Great-tailed grackles are now almost exclusively associated with human modified environments Wehtje (2003), and when planning study sites, we initially wanted to compare forest versus urban grackle populations. However, we are unable to find a population that exclusively exists in forests (based on eBird.org data, Logan, pers. obs.). In another article produced from the same preregistration, Logan et al. (2020), as the current article, we investigated the role of increased habitat availability in geographic range expansions by comparing rapidly expanding great-tailed grackles with their closest relative that is not rapidly expanding its range, boat-tailed grackles (*Q. major*) (Summers et al., 2023). We predicted that great-tailed grackles expanded their range because suitable habitat (i.e., human modified environments) increased (prediction 1 alternative 1 in the preregistration). Results showed that, between 1979 and 2019, great-tailed grackles increased their habitat breadth to include more urban, arid environments. In contrast, boat-tailed grackles moved into new suitable habitat that was made available by climate change. These results support the possibility that flexibility played a role in the ability to increase habitat breadth. We are currently conducting a behavioral flexibility experiment in boat-tailed grackles to determine whether they are less flexible than great-tailed grackles, which would further support the hypothesis that flexibility was involved in the great-tailed grackle rapid range expansion (in the same preregistration as the current study: Logan et al., 2020). Unfortunately, we discovered in our first boat-tailed grackle field season in 2022 that they do not do well in captivity. Consequently, we will not continue the aviary tests in this species. Therefore, we only have comparable data from the aviary tests for two (reversal), four (multiaccess box), and five (persistence and exploration) individuals. Although the boat-tailed grackle sample size is too small to arrive at robust conclusions, we analyze their data here to give an indication of useful directions for future research. We find that boat-tailed grackles have a **similar flexibility average** as both populations of great-tailed grackles; and boat-tailed grackles are **less innovative** and **less persistent** than both great-tailed grackle populations. Boat-tailed grackles are **less exploratory** than Tempe grackles, while having **similar levels of exploration** as Woodland grackles (see model outputs in Supplementary Material 4). This suggests that we might not find differences in flexibility between the two species. However, we are currently conducting reversal learning experiments in the wild in both species to determine whether this is a robust result Logan et al. (2022).

The ability of great-tailed grackles to move into new habitats might be a species specific ability that has been ongoing for many years, and could be linked to the high levels of flexibility in this species being relatively fixed (Wright et al., 2010). Great-tailed grackles are flexible on the reversal learning task and are perhaps at their upper limit uniformly across their range. With an average reversal learning speed of 74 trials (using the data in the current article), great-tailed grackles are as flexible as great (*Parus major*) and blue (*Cyanistes caeruleus*) tits (average 59 trials, Morand-Ferron et al., 2022) and three species of Darwin's finches (*Camarhynchus parvulus*, *C. pallida*, and *Geospiza fortis*, average 89 trials); and more flexible than Pinyon jays (average 155 trials), Clark's nutcrackers (average 143 trials), California scrub jays (average 191 trials), pigeons (average 168 trials) (data reported in Tebbich et al., 2010; but not in the original articles Bond et al., 2007; and Lissek et al., 2002), and mice (average approximately 150 trials, Laughlin et al., 2011). Perhaps great-tailed grackles maintain a high level of flexibility across their range in response to daily changes in their local environment (e.g., the changing schedules of cafes with outdoor seating areas and garbage pick up times, Rodrigo et al., 2021), rather than specifically in response to larger changes that might occur less frequently (e.g., traveling farther to exploit new foraging opportunities or moving to a new area).

Another alternative is that we measured the edge population too long after their initial establishment, during which time they potentially exhibited more flexibility for their initial adaptation phase to the new area (Wright et al., 2010). Though it seems that this population is still becoming established, in that they are not found at the Woodland trap site year-round and some individuals at the Sacramento trap site also disappear and reappear for parts of the year. 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 to 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. In this case, because differences in persistence were found, this trait could have different timing in the process of establishing in a new location (i.e., be required for longer). Great-tailed grackles occur more irregularly in areas further north of our edge site, and flexibility might be higher in more northern individuals from areas where no stable populations are yet established. Because the more northern populations are still small and ephemeral, to obtain our minimum sample sizes, a different and more geographically expansive experimental approach would be necessary. Future efforts could focus on a broader geographic area across Washington or Oregon for capturing these individuals to measure flexibility and other behaviors to add important information to our understanding of the relationship between variation in behavior and the ability of species to expand their range. However, evidence from experimental evolution suggests that, even after 30 generations there is no change in exploration of a novel environment or other behaviors (aggression, social grooming, courtship, and orientation) 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). This means it is likely that we would have detected population differences if such differences were linked with adapting to a new environment.

While great-tailed grackles are not considered an invasive species because they expanded their range without direct human assistance, comparing them with invasive species is useful because the dynamics after the introduction stage should be similar (i.e., establishing in a new area and spreading out from there) (Chapple et al., 2012). Note that wild great-tailed grackles were caught from north of Rio de la Antigua, Mexico by the Aztec emperor, Auitzotl (1486-1502), and introduced approximately 370 km inland to the Valley of Mexico

(Tenochititlan & Tlatelolco) where they reproduced and spread (Haemig, 2011, 2012; Haemig, 2014). By 1577, they spread at least 100 km including back to their native range (Haemig, 2011). This indicates that great-tailed grackles had already spread this far north by themselves before the introduction at a parallel latitude, and that they continued their spread without the help of human-facilitated introductions.

In conclusion, rather than flexibility being higher on average in an edge population of a species undergoing a rapid geographic range expansion, as is widely hypothesized, we found that a higher variance in flexibility and higher average in persistence were the key behavioral traits associated with the great-tailed grackle's edge population in comparison with an older population closer to the original range. This calls into question the importance of several traits that are hypothesized to be involved in such an expansion. The term "behavioral flexibility" is defined and measured in a variety of ways in the literature (or it is not defined at all) (Audet & Lefebvre, 2017). For example, the detour task (individuals must walk around a transparent barrier to access a food reward) is sometimes considered a test of flexibility (e.g., Troisi et al., 2020), sometimes a test of self control (MacLean et al., 2014; e.g., Isaksson et al., 2018; Knolle et al., 2019), and sometimes a test of both (e.g., Storks & Leal, 2020). However, theoretically and empirically it measures a trait that is not, and is not related to, flexibility or self control, but rather a different trait: motor inhibition (Beran, 2015; Logan et al., 2021). We argue that calling many types of traits "flexibility" without proper (or sometimes any) theoretical justification and without validating methods is detrimental because it confounds our ability to answer questions about the broader significance of flexibility and how it is genuinely involved in large scale changes (Audet & Lefebvre, 2017; Logan et al., 2017; Mikhalevich et al., 2017). Our research program shows the value of clearly defining terms for behavioral traits, validating the methods intended to measure those traits, and understanding how certain traits relate to each other (causally if possible) before attempting to answer broader cross population questions.

DATA, SCRIPT, AND CODE AVAILABILITY

Data, scripts, and code are available online at the Knowledge Network for Biocomplexity's data repository: https://doi.org/10.5063/F10C4T7T (Logan et al., 2023c). The scripts and code are in the Rmd file, which is also available at https://github.com/corinalogan/grackles/blob/master/Files/Preregistrations/gxpopbehaviorhabitatq1.Rmd

ETHICS

This research was 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)
6) RegionalSan (access permit number AP 2021-01)

AUTHOR CONTRIBUTIONS

Logan: Hypothesis development, data collection, data analysis and interpretation, write up, revising/editing, materials/funding.

McCune: Method development, data collection, data analysis and interpretation, revising/editing.

LeGrande-Rolls: Data collection, revising/editing.

Marfori: Data collection, revising/editing.

Hubbard: Data collection, revising/editing.

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

FUNDING

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

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 used to be on the Managing Board at PCI Ecology (2018-2022).

ACKNOWLEDGEMENTS

We thank our PCI Ecology recommender, Esther Sebastián González, and reviewers, Francois-Xavier Dechaume-Moncharmont, Pizza Ka Yee Chow, and one anonymous reviewer of the post-study article for their useful feedback, which greatly improved the manuscript. We also thank our pre-study manuscript PCI Ecology recommender, Esther Sebastián González, and reviewers Caroline Marie Jeanne Yvonne Nieberding, Pizza Ka Yee Chow, Tim Parker, and 1 anonymous reviewer for their helpful feedback before the study began. We are grateful to 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; Aaron Blackwell and Ken Kosik for being the UCSB sponsors of the Cooperation Agreement with the Max Planck Institute for Evolutionary Anthropology; Alexis Breen and Vincent Kiepsch for video coding; Ann Brice, Woodland-Davis Clean Water Agency, RegionalSan, and Conaway Ranch for hosting the research on their land; and Dr. Rhonda Oates and the vet team at UC Davis for veterinary consultations.

SUPPLEMENTARY MATERIAL 1: Sample size rationale

We summarize the minimum sample sizes and their associated detection limits in Table SM1, which allows us to determine whether populations are different from each other (detailed in the Analysis section for each experiment).

Table SM1. 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 SM2.1 and SM2.2	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 SM2.1 and SM2.2)	15
Multiaccess box	Number of loci solved	Binomial	Differences of 1.0 loci are likely to be detected (Supplementary Material SM2.1)	15
Exploration	Latency to approach novel object	Gamma-Poisson	Differences of at least 1407 sec are likely to be detected for n=14/site and 824 sec for n=20/site (Supplementary Material SM2.2)	14
Persistence	Proportion of trials participated in	Normal	Difference of at least 0.1 in the proportion of trials participated in (Supplementary Material SM2.3)	15

SUPPLEMENTARY MATERIAL 2: Simulations for power analyses

Hypothesis-specific mathematical model

Following procedures in McElreath (2020a), we constructed a **hypothesis-appropriate mathematical model** for each of the response variables that examines differences in the response variable between sites (each site represents a grackle population). Except for the flexibility model, which was modeled on data from Santa Barbara where no flexibility manipulation occurred (thus no Treatment variable), these models take the form of:

$y_i \sim \alpha[\text{site}] + \beta * treatment$,

where y_i is the response variable (flexibility, innovation, exploration, or persistence). There is one intercept, $\alpha[\text{site}]$, per site, β is the expected amount of change in the response variable for each *treatment* (flexibility manipulated or not). We estimate the site's average and standard deviation of the response variable. The flexibility model only has the $\alpha[\text{site}]$ term.

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 then performed pairwise contrasts to determine at what point we can detect differences between sites by manipulating sample size, 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 used a Bayesian approach, therefore comparisons are based on samples from the posterior distribution. We drew 2,000 samples from the posterior distribution, where each sample had an estimated mean for each population. For the first contrast, within each sample, we subtracted the estimated mean of the edge population from the estimated mean of the core population. For the second contrast, we subtracted the estimated mean of the edge population from the estimated mean of the middle population. For the third contrast, we subtracted the estimated mean of the middle population from the estimated mean of the core population. We then had samples of differences between all of the pairs of sites, which we use to assess whether any site is systematically larger or smaller than the others. We determined 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.

Flexibility analysis

Power analyses: We also used the simulations to estimate our ability to detect differences in ϕ and λ between sites based on extracting samples from the posterior distribution. We ran two different sets of simulations: we first sampled between 9 and 24 birds from populations with pre-specified ϕ and λ means to determine the minimum sample size required to detect whether two populations are different. This set of simulations showed 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 SM2.1). The second set of simulations recreated choices for 20 birds per population across initial learning and reversal trials from which we estimate their ϕ and λ. We simulated 20 birds per population because this number is above the threshold we detected in the first set of simulations and it appears a feasible sample size. We expected that the noise in the probabilistic choices of individuals might reduce the differences that can be detected compared to the first simulation where ϕ and λ are assumed to be exactly known for each individual. This second 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 ϕ is 0.01 or greater and/or if the difference in their λ is 3 or greater, based on data from 20 simulated individuals per site (Figure SM2.2). It appears that there is more variability in the λ estimates for each bird based on their choices, meaning that with the learning model, which estimates λ from the choices, the differences between sites have to be larger (than if we were able to infer λ directly) to be reliably detected. Given that we have to infer ϕ and λ from the choices, the power curves in Figure SM2.1 are more reliable than those in Figure SM2.2.

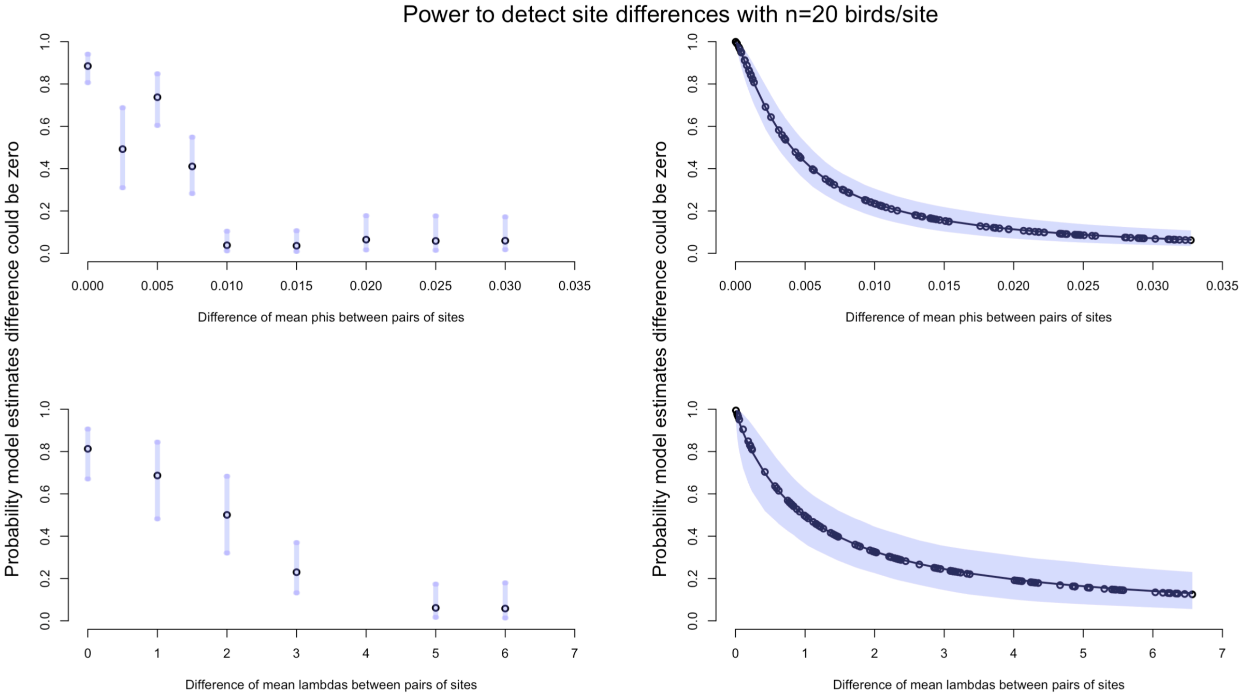

Figure SM2.1. How small of a site difference in ϕ and λ can we detect? The probability that the model estimates that the difference shown on the x axis is zero, meaning that the model assumes that it is possible that these two estimates come from a population with the same ϕ or λ. Each point is the mean ϕ or mean λ from one site minus the mean ϕ or mean λ from another site (calculated from 20 individuals per site) for all pairwise comparisons for all 32 simulated sites (for a total of 496 pairwise comparisons). Left panels: error bars=89% compatibility intervals. Right panels: shaded areas=97% prediction intervals.

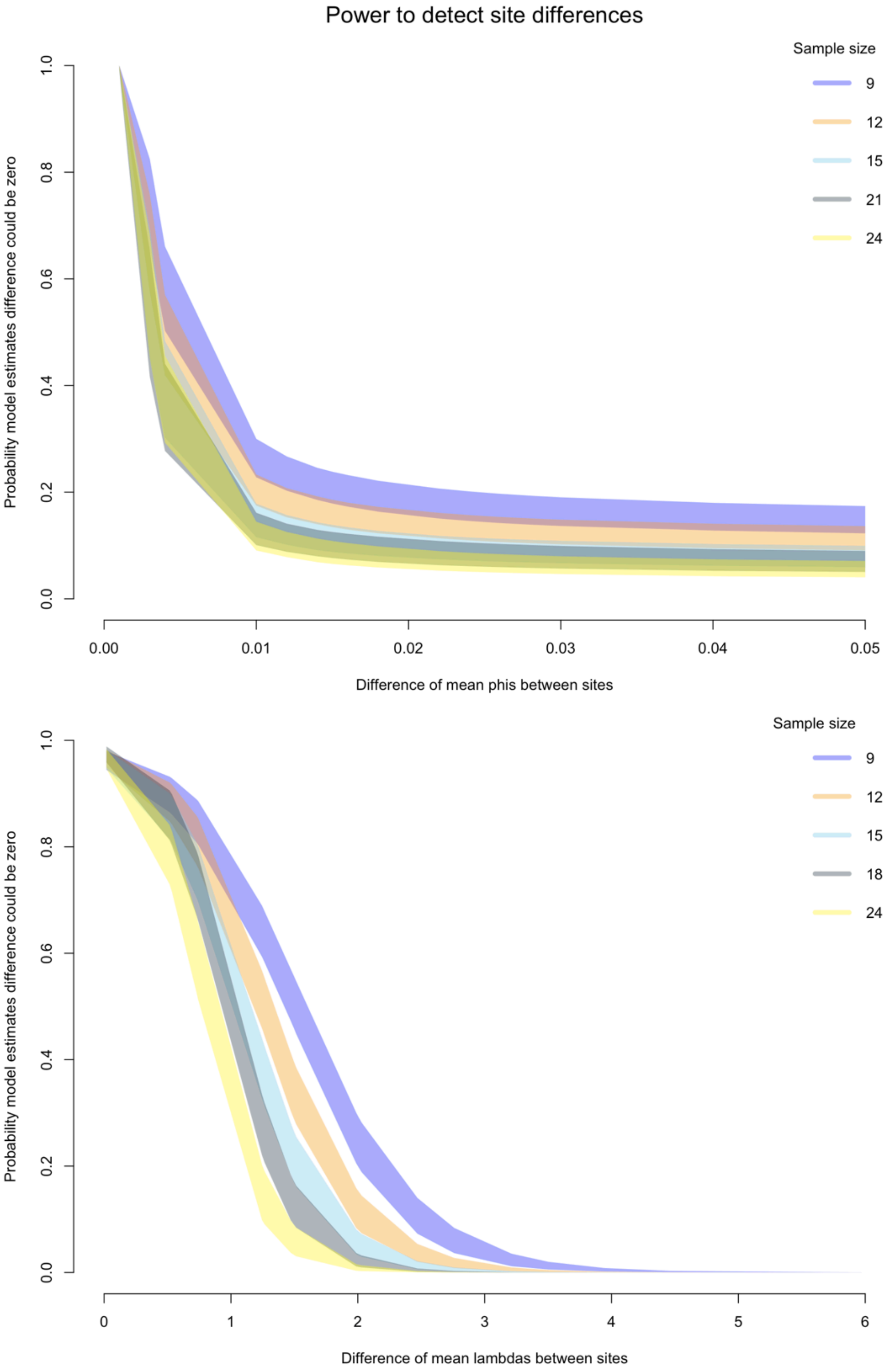
Power to detect site differences
Sample size
9
12
15
21
24
Probability model estimates difference could be zero
1.0
0.8
0.6
0.4
0.2
0.0
0.00
0.01
0.02
0.03
0.04
0.05
Difference of mean phis between sites
Sample size
9
12
15
18
24
Probability model estimates difference could be zero
1.0
0.8
0.6
0.4
0.2
0.0
0
1
2
3
4
5
6
Difference of mean lambdas between sites

Figure SM2.2. 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 ϕ or λ. The x-axis is the mean ϕ or mean λ from one site minus the mean ϕ or mean λ 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.

Innovation analysis

After building the model (see Methods), we then ran the **mathematical model** and performed pairwise contrasts and determined that we are able to detect differences between sites with a sample size of 15 at each site if the average number of loci solved differs by 1.0 loci or more, the standard deviation is generally a maximum of 0.1 at each site, and the flexibility manipulated individuals are slightly (or much) better than the non-manipulated individuals (Table SM2.1). For a sample size of 20 at each site, we are able to detect site differences if the average number of loci solved differs by 0.8 of a locus or more, the standard deviation is generally a maximum of 0.1 at each site, and the flexibility manipulated individuals are much better than the non-manipulated individuals (Table SM2.1). Note: the Arizona sample size is 12 for the multiaccess log and 17 on a similar multiaccess box.

Table SM2.1 Sample size is the number of individuals per site multiplied by two sites (e.g., n=15 per site indicates that 30 individuals were involved in this simulation), settings combination is the combination of settings for site differences and manipulation effects used for a given simulation run, site differences are the simulated differences between the two site means in the proportion of loci solved, manipulation effect is the simulated difference in the proportion of loci solved between the flexibility manipulated and non manipulated birds, X/10 crosses zero is the number of times out of the 10 repetitions for this setting combination in which the contrast between sites crosses zero (if it did cross zero, then we did not detect site differences).

Sample size	Settings combination	Site difference	Manipulation effect	X/10 crosses zero
15	1	0.10	0.00	6
15	2	0.10	0.10	7
15	3	0.10	0.25	8
15	4	0.15	0.00	5
15	5	0.15	0.10	7
15	6	0.15	0.25	4
15	7	0.20	0.00	3
15	8	0.20	0.10	3
15	9	0.20	0.25	2
15	10	0.25	0.00	1
15	11	0.25	0.10	0
15	12	0.25	0.25	0
20	1	0.10	0.00	7
20	2	0.10	0.10	5
20	3	0.10	0.25	6
20	4	0.15	0.00	4
20	5	0.15	0.10	5
20	6	0.15	0.25	4
20	7	0.20	0.00	1
20	8	0.20	0.10	1
20	9	0.20	0.25	0
20	10	0.25	0.00	1
20	11	0.25	0.10	0
20	12	0.25	0.25	0

Because the mean and the variance are linked in the binomial distribution, and because the variance simulations in the flexibility analysis showed that we are not able to robustly detect differences in variance between sites, we 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.

Exploration analysis

After building the model (see Methods), we then ran the **mathematical model** and performed pairwise contrasts and determined that we are be able to detect differences between sites with a potential sample size of 14 at each site if the average latency to approach the novel environment differs by at least 1407 sec between sites and 824 sec for a sample size of 20 at each site (Table SM2.2). We kept the shape of the curve (which can be thought of as similar to a standard deviation or the variance) the same across sites because we did not think this assumption would change across populations (i.e., there could 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).

Table SM2.2 Sample size is the number of individuals per site multiplied by two sites (e.g., n=14 per site indicates that 28 individuals were involved in this simulation), settings combination is the combination of settings for site differences and manipulation effects used for a given simulation run, site differences are the simulated differences between the two site means of latency to approach a novel environment, manipulation effect is the simulated difference in the latency between the flexibility manipulated and non manipulated birds, X/10 crosses zero is the number of times out of the 10 repetitions for this setting combination in which the contrast between sites crosses zero (if it did cross zero, then we did not detect site differences).

Sample size	Settings combination	Site difference	Manipulation effect	X/10 crosses zero
14	1	0.0	0.00	9
14	2	0.0	0.10	10
14	3	0.0	0.25	9
14	4	2.0	0.00	4
14	5	2.0	0.10	1
14	6	2.0	0.25	2
14	7	2.5	0.00	0
14	8	2.5	0.10	3
14	9	2.5	0.25	0
14	10	3.0	0.00	1
14	11	3.0	0.10	0
14	12	3.0	0.25	0
20	1	0.0	0.00	10
20	2	0.0	0.10	10
20	3	0.0	0.25	8
20	4	2.0	0.00	2
20	5	2.0	0.10	1
20	6	2.0	0.25	3
20	7	2.5	0.00	0
20	8	2.5	0.10	0
20	9	2.5	0.25	0
20	10	3.0	0.00	0
20	11	3.0	0.10	0
20	12	3.0	0.25	0

Persistence analysis

After building the model (see Methods), we then ran the **mathematical model** and performed pairwise contrasts and determined that we are be able to detect differences between sites with a potential sample size of 15 or 20 per site if the average proportion of trials participated in differs by at least 0.1 if there are not strong effects from the flexibility manipulation and at least 0.2 if there are strong flexibility manipulation effects, and the standard deviation is 0.1 at each site (Table SM2.3).

Table SM2.3 Sample size is the number of individuals per site multiplied by two sites (e.g., n=15 per site indicates that 30 individuals were involved in this simulation), settings combination is the combination of settings for site differences and manipulation effects used for a given simulation run, site differences are the simulated differences between the two site means in the proportion of trials participated in, manipulation effect is the simulated difference in the proportion of trials participated in between the flexibility manipulated and non manipulated birds, X/10 crosses zero is the number of times out of the 10 repetitions for this setting combination in which the contrast between sites crosses zero (if it did cross zero, then we did not detect site differences).

Sample size	Settings combination	Site difference	Manipulation effect	X/10 crosses zero
14	1	0.0	0.00	9
14	2	0.0	0.10	10
14	3	0.0	0.25	9
14	4	2.0	0.00	4
14	5	2.0	0.10	1
14	6	2.0	0.25	2
14	7	2.5	0.00	0
14	8	2.5	0.10	3
14	9	2.5	0.25	0
14	10	3.0	0.00	1
14	11	3.0	0.10	0
14	12	3.0	0.25	0
20	1	0.0	0.00	10
20	2	0.0	0.10	10
20	3	0.0	0.25	8
20	4	2.0	0.00	2
20	5	2.0	0.10	1
20	6	2.0	0.25	3
20	7	2.5	0.00	0
20	8	2.5	0.10	0
20	9	2.5	0.25	0
20	10	3.0	0.00	0
20	11	3.0	0.10	0
20	12	3.0	0.25	0

SUPPLEMENTARY MATERIAL 3: Interobserver reliability of dependent variables

To determine whether experimenters coded the dependent variables in a repeatable way, hypothesis-blind video coders were first 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 indicated that the two coders (the experimenter and the video coder) agreed with each other to a high degree (kappa: Landis & Koch, 1977; ICC: Hutcheon et al., 2010). After passing training, the video coders coded 20% of the videos for each experiment (except for exploration for which 15% of the videos were coded due to an unexpectedly high sample size for this assay). The kappa and ICC were 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 was 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 was 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 et al., 2019b) 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 15% of the videos) were as follows:

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

- Reversal learning (5/19 birds): correct choice unweighted Cohen's Kappa=1.00 (confidence boundaries=0.99-1.00, n=707 data points)
- Multiaccess box (5/23 birds): correct choice unweighted Cohen's Kappa=0.92 (confidence boundaries=0.81-1.00, n=63 data points)
- Multiaccess box (5/23 birds): 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 (5/34 birds): latency to land on the ground unweighted Cohen's Kappa=0.998 (confidence boundaries=0.997-0.999, n=32 data points)

SUPPLEMENTARY MATERIAL 4: Boat-tailed grackle model outputs

Table SM4. Results for the comparison between the boat-tailed grackle (BTGR) population in Lake Placid and Venus, Florida and the great-tailed grackle populations in Tempe, Arizona and Woodland, California. Contrasts (indicated by “diff”) between populations show whether there was a difference (compatibility interval does not cross zero) or not (compatibility interval crosses zero) for that pair of populations. Populations are labeled as follows: 1=boat-tailed grackles (BTGR), 2=Woodland great-tailed grackles, 3=Tempe great-tailed grackles (e.g., diff_12 means that BTGR and Woodland are being compared).

	Mean	Standard deviation	Lower 89 percentile compatibility interval (5.5%)	Upper 89 percentile compatibility interval (94.5%)
FLEXIBILITY	NA	NA	NA	NA
BTGR phi	0.09	0.03	0.07	0.11
diff_12 phi	0.02	0.02	-0.01	0.06
diff_13 phi	0.02	0.02	-0.01	0.06
BTGR lambda	5.17	1.16	4.44	5.90
diff_12 lambda	0.07	1.11	-1.52	2.00
diff_13 lambda	0.32	1.15	-1.20	2.39
	NA	NA	NA	NA
INNOVATIVENESS	NA	NA	NA	NA
BTGR	-0.47	0.55	-1.35	0.41
Woodland	1.20	0.24	0.82	1.58
Tempe	1.27	0.43	0.63	1.99
diff_12	-0.37	0.13	-0.58	-0.15
diff_13	-0.38	0.15	-0.61	0.14
	NA	NA	NA	NA
PERSISTENCE	NA	NA	NA	NA
BTGR	0.82	0.09	0.67	0.98
Woodland	1.28	0.04	1.21	1.36
Tempe	1.13	0.05	1.06	1.21
diff_12	-0.09	0.02	-0.12	-0.06
diff_13	-0.06	0.02	-0.10	-0.03
	NA	NA	NA	NA
EXPLORATION	NA	NA	NA	NA
BTGR	7.93	0.35	7.42	8.52
Woodland	7.55	0.13	7.35	7.76
Tempe	7.34	0.20	7.02	7.65
diff_12	1042.09	1195.19	-292.29	2884.97
diff_13	2946.79	1164.04	1657.26	4985.57

SUPPLEMENTARY MATERIAL 5: Protocols

PROTOCOLS for Flexibility Experiment (reversal learning) Innovativeness Experiment (multiaccess log) Exploration Assay (novel object/environment)

Counterbalancing order of experiments and the first rewarded color in reversal learning

Table SM5.1. Counterbalancing the first rewarded color (light gray=1 or dark gray=2) for the reversal learning experiment, the order of experiments (reversal learning=1 and multiaccess log=2), and which locus they were trained to demonstrate for the learning mechanism experiment (see McCune et al., 2019b for details); we will train half of the demonstrators in each batch on one solving method on the log apparatus (Bup) and the other half of the demonstrators in each batch on one solving method on the plank apparatus (Vflap). One batch = 8 birds tested at one time. Bird number refers to the number of the aviary they are housed in (1-8). Random numbers were generated using https://www.random.org. NOTE: the Woodland population experiences the plank apparatus first, then the log apparatus afterward. The population in the core of the range experiences the reverse. *Piña was initially assigned the Log apparatus for demonstrator training, but was then switched to the Plank apparatus after 2 days of training on Bup because we needed to release her quickly and Bup is not quick to learn for grackles; therefore, we randomly chose one Plank demonstrator from batches 2 and 3 and switched them to a Log demonstrator to equalize counterbalancing (batch 3, bird 6, random.org). NOTE: On 9 Mar after 3 weeks of unsuccessfully training Tembleque on Bup, we switched to training him on Bdown to see if it will be easier for grackles to learn. If so, then we would change all birds assigned to Bup to Bdown instead. It was not easier to learn, therefore we stopped training demonstrators on the log and removed it from the social learning experiment. After Tembleque, all birds were only trained on the plank apparatus.

Batch	Bird	Name	First experiment	First rewarded color	Demonstrate	Batch	Bird	Name	First experiment	First rewarded color	Demonstrate
1	1	Xango	Multiaccess	Light gray	Plank	4	1	Bacmut bacni	Reversal	Light gray	Plank
1	2	Flan	Reversal	Dark gray	Plank	4	2	Changa	Reversal	Light gray	Plank
1	3	Camote	Multiaccess	Dark gray	Plank	4	3	Cutuy	Multiaccess	Light gray	Plank
1	4	Tembleque	Multiaccess	Dark gray	Log	4	4	Xunub	Reversal	Dark gray	Plank
1	5	Polvorones	Multiaccess	Light gray	Plank	4	5	Urraca	Multiaccess	Dark gray	Plank
1	6	Dulce de Leche	Reversal	Light gray	Log	4	6	Cocinera	Multiaccess	Light gray	Plank
1	7	Alegria	Reversal	Dark gray	Log	4	7	Tzanatl preciosa	Multiaccess	Dark gray	Plank
1	8	Helado	Reversal	Light gray	Log	4	8	Quiscalus	Reversal	Dark gray	Plank
2	1	Zapote Negro	Multiaccess	Light gray	Plank	5	1	Verbena	Reversal	Light gray	NA
2	2	Piña	Multiaccess	Light gray	*Plank	5	2	Cocoplum	Reversal	Light gray	NA
2	3	-	Reversal	Dark gray	Plank	5	3		Multiaccess	Dark gray	NA
2	4	Carlota	Reversal	Light gray	Plank	5	4		Multiaccess	Light gray	NA
2	5	-	Reversal	Light gray	Plank	5	5		Reversal	Light gray	NA
2	6	Buñuelo	Multiaccess	Dark gray	Plank	5	6		Reversal	Dark gray	NA
2	7	-	Multiaccess	Dark gray	Plank	5	7		Multiaccess	Dark gray	NA
2	8	-	Reversal	Dark gray	Plank	5	8		Multiaccess	Dark gray	NA
3	1	Ak’xi’	Reversal	Dark gray	Plank	6	1		Reversal	Light gray	NA
3	2	Kau	Multiaccess	Dark gray	Plank	6	2		Multiaccess	Light gray	NA
3	3	Galandra	Multiaccess	Dark gray	Plank	6	3		Reversal	Light gray	NA
3	4	Kel	Multiaccess	Dark gray	Plank	6	4		Multiaccess	Dark gray	NA
3	5	Cuervo	Reversal	Light gray	Plank	6	5		Multiaccess	Light gray	NA
3	6	-	Multiaccess	Light gray	Plank	6	6		Reversal	Dark gray	NA
3	7	Wachil	Reversal	Light gray	Plank	6	7		Reversal	Dark gray	NA
3	8	Talingo	Reversal	Light gray	Plank	6	8		Multiaccess	Dark gray	NA

Table SM5.2. Counterbalancing the first exploration assay, environment (env) or object (obj), for those grackles who received both assays. The Arizona exploration and boldness data, the results of which will determine whether we can use only one exploration assay, was not done being analyzed by the time the Woodland field season started in January 2021. Therefore, we continued with both assays until the Arizona results were finalized. The order for each bird was randomized using random numbers generated by https://www.random.org (1=environment first, 2=object first). For those birds who experienced both environment and object assays, they were conducted on consecutive days. *=this bird did not complete experiments and was therefore replaced in that batch and aviary.

Batch	Bird	Name	Exploration assay order (1st, 2nd)
1	1	Xango	Env, Obj
1	2	Flan	Obj, Env
1	3	*Churro	Obj, Env
1	4	Tembleque	Obj, Env
1	5	Polvorones	Env, Obj
1	6	*Chocolate	Obj, Env
1	7	Alegria	Env, Obj
1	8	*Sopapilla	Env, Obj

FLEXIBILITY: reversal learning (experimental design after Logan, 2016b)

Apparatus: color tubes: 3x light gray tubes, 3x dark gray tubes, 6x yellow tubes

Yellow tube training

Summary: Get the bird used to searching for food that is out of sight inside a tube. First, habituate the bird to the yellow tube by placing it in their food dish at least one day before testing. Then, start yellow tube training where they learn to search for food hidden inside a yellow tube. If, after starting yellow tube training they still appear scared of the tube, keep putting it in their food bowl overnight until they are habituated.

Habituation to yellow tube: leave yellow tube in food dish overnight. Note when the yellow tube was left in the bird's food dish overnight in the Notes section of the first (or next) trial of Training: Yellow Tube

Training: Yellow Tube

Training trials are not video recorded

Data sheet: data_xpop > tab: data_yellowtubetraining

Description: Use a yellow tube to train them to search for hidden food. Place the baited (with food inside at the back of the tube) tube on the table or ground (and move all other objects away from the testing area) so the bird can see the food (place food on the lip/tube opening and on the table or ground around the front of the tube). Wait for them to eat the food. Repeat while placing the tube at various places on the table or ground (to avoid associating food with a location), while gradually turning the tube so the food is not visible. In the beginning, food may be added to tube in view of the bird. Record the progression of whether food and tube were visible or not visible to the bird on each trial in the Notes column. To count toward criterion, the experimenter must place the food inside the tube out of view of the bird and then the tube must not face the bird so the bird must rely solely on the knowledge that they have to search for food that is not visible.

- How to score the "correct choice" column in the data sheet:
  - 0 = eat from around the tube but not inside it
  - 1 = eat the food from inside the tube
  - -1 = they do not participate (they don't eat any food)

  - NOTE: when scoring an individual session (i.e., 1 session = 1 row in the data sheet) rather than an individual trial (because the bird is not yet participatory enough for trial level data), score each session according to the highest number they achieved across the whole session. For example, if there were 5 trials in the session and the bird took only visible food and not nonvisible food, then score the whole session as 0.

- Once the bird is readily participating and obtaining the food when it is not visible, they must pass this criterion: successfully obtain the food from the tube when it is not visible on 5 consecutive trials within a session or across sessions in one day (i.e., score=1, indicate these are trials that contribute toward meeting criterion in the column "Criterion: successfully obtain the food from the tube when it is not visible in 5 consecutive trials on the same day").

Testing Training: food on ground/table

Training trials are not video recorded

Data sheet: data_xpop > tab: data_yellowtubetraining

In Notes column say: "no yellow tube was present"

If a bird is not participating in yellow tube trials by not coming down to search for food within 5 min, remove the yellow tube and place a piece of food on the ground or on the table for up to 5 min. If they do not eat it, remove the food and try again later. This trains them to come down and eat within 5 min, otherwise the food is removed and they won't have access to food again until the next session

Training: no color preference

All color preference training trials are video recorded (put camera at rear of aviary to film the bait on the tubes)

Data sheet: data_xpop > tab: data_nocolorpref

Obtain pseudorandomized order for which color is on which side: RandomizationsReversalLearning

Video file naming convention: A035P- 2018-11-04 Reversal Training No Color Preference S2 T4

Description: The purpose of this training is to remove any potential initial color preference to ensure the bird attends to the functional properties of the task when the experiment begins. Birds are given 10 color preference trials for light gray and dark gray tubes by presenting both tubes (one of each color) on the table at the same time and in a pseudorandomized order for side (alternating sides for the first two trials of a 10-trial set, presenting the same tube on the same side up to two times consecutively thereafter). The tube openings are taped shut.

1) Place tubes on the table (or floor - and move all other objects away from the testing area) at the same time spaced approximately 30 cm apart and with the taped tube openings facing the back wall of the aviary
2) Place two pieces of food (goldfish crackers, peanuts, or maintenance diet) on top of both tubes at the same time (on top of the wooden piece at the back of the tube), then two pieces at the front of both tubes at the same time (on top of the wood, in front of the tube opening).
3) Record the first tube from which a bird eats food (this is considered its color choice). Allow the bird to eat all of the food from both tubes before starting the next trial.
4) If an obvious color preference develops as habituation trials progress (i.e., if the bird approaches the same color first 9 or 10 times out of the most recent 10 trials, which is statistically significant according to a binomial test), more food is placed on the least preferred color to reduce the preference. If a bird chooses the same color 4 times in a row, start to load more food on the other color.
5) Repeat 10-trial sessions until the bird shows no color preference (the 10 trials can occur across sessions and/or days).
6) If bird doesn't come down within 5 minutes, end session and try again in the next session.

7) **Habituation as needed:** If a bird is hesitant to approach the tubes in their first 10 trials, put one light gray and one dark gray tube (both with openings taped over) in their food dish overnight until they are habituated. Ensure the tube openings are taped over so they do not associate getting the food out of the inside of the tube of one color more than the other.

How to score the "correct choice" column in the data sheet: - 1 = ate food first from the rewarded color (both colors are rewarded here, but use their first rewarded color in the test for coding purposes) - 0 = ate food first from the non-rewarded color (both colors are rewarded here, but use their first rewarded color in the test for coding purposes) - -1 = they did not eat food from either tube. This trial is incomplete and is re-conducted until the bird eats the food

Criterion to pass: choose one color 8 or fewer times out of the most recent 10 trials (counting in a 1 trial sliding window), indicating no color preference. Move the bird on to the Test.

Test
All trials are video recorded

Data sheet: data_xpop > tab: data_reverse

Obtain pseudorandomized order for which color is on which side: RandomizationsReversalLearning

Video file naming conventions: A043KR 2018-12-22 Reversal 0 S16 T35

Initial discrimination
Description: One light gray and one dark gray tube are "placed at opposite ends of a table (or on the floor - and move all other objects away from the testing area) with the tube openings facing the side walls so the bird could not see which tube contained the food. Tubes were pseudorandomized for side and the left tube was always placed first, followed by the right to avoid behavioral cueing. Pseudorandomization consisted of alternating location for the first two trials of a session and then keeping the same color on the same side for at most 2 consecutive trials thereafter. Each trial consisted of placing the tubes on the table or floor, and then the bird had the opportunity to choose one tube by looking into it (and eating from it if it chose the rewarded tube). Once the bird chose, the trial ended by removing the tubes" (Logan 2016 PeerJ). To avoid behavioral cueing, always enter the aviary to set up the experiment, then turn to the right when leaving, turn to the right after re-baiting, and re-enter the aviary.

1) Prepare datasheet with 10 or more trials (enter all info except for StartTime and CorrectChoice). To fill in OptionOnLeft, open the "Randomized Sessions" datasheet. Follow instructions in this datasheet for retrieving a list of r/n's for a session (r = rewarded color, n = non-rewarded color). Note that if a session includes the end of one set of 10 trials of randomization and the beginning of another set of 10 trials of randomization, make sure the pseudorandomization rules aren't broken by rearranging the first couple of trials of the next randomization if necessary.
2) Record the time into the datasheet for at least the first trial. Record start times if possible for later trials, but not necessary if the grackle is moving quickly.
3) Bait the rewarded tube (make sure no grackle in any of the aviaries can see what you are doing). Hold tubes with openings facing away from you and fingers covering the tube openings. Tilt the tubes slightly backwards so the food does not fall out or make noise.
4) Go into aviary and place the left tube, then the right tube so they are equidistant from edges (~6 inches from each edge of the table or from the side walls if placed on the floor). Make sure the food does not make noise inside the tube as you set tubes down. Leave the aviary by turning to the right. Watch grackle from outside the aviary.
5) A choice is recorded if they bend their head and/or body down to look inside a tube (this was updated on 10 Oct 2018. Previously, a choice was counted if they passed an imaginary line perpendicular to the opening of the tube. However, they can not actually see the food unless they bend their head down). NOTE (23 Mar 2021): make sure that the tubes are sitting flat on the ground.

How to score the "correct choice" column in the data sheet:

- 1 = chose the rewarded color and had access to the food reward (regardless of whether they chose to eat it)
- 0 = chose the non-rewarded color
- -1 = did not make a choice. This trial is incomplete and is re-conducted until the bird makes a choice.

6) Birds are only allowed to look into one tube per trial. If they try to look in the other tube after they already made a choice (looked inside a tube), interrupt them before they can see inside the other tube, and reset the trial. They may look inside their chosen tube, retrieve the food (if they choose the rewarded color), walk around the tube, etc. If a grackle wants to drink after a trial, let them finish before entering the aviary to start the next trial.
7) Rebait (or pretend to rebait if food was not eaten) and conduct the next trial.
8) If a bird chooses the same side on 4 consecutive trials, they might have a side bias, in which case, stop the current random numbers for side and start putting the rewarded color on the non-preferred side as much as possible while still following the pseudorandomization rules (above in italics). Also, if they usually start from a particular perch, angle the table so it is parallel to that perch. Only give them a maximum of 10 trials per session if they have a side bias.
9) If the grackle has not made a choice in 2-3 minutes (general rule), you can place a small food piece (usually smaller than the piece of food inside the tube, but can be bigger or multiple pieces as long as they make a choice after eating it and do not just eat this piece of food without making a choice) equidistant between the tubes to entice them to participate. If they come down and only eat the bait and do not make a choice, then do not bait again until after they make a choice. If the grackle has not made a choice in 3-5 minutes, end the session and try again later. This helps them learn to work faster. Some individuals work really slowly and 5 minutes maximum would never work for them so, for these individuals, work at their pace if you have time.
10) Session = a continuous opportunity for a bird to participate in as many trials as they are interested in participating in, which begins when they are offered the opportunity and ends when their motivation to participate wanes or they complete enough trials to complete a chunk of the experiment (generally ~20 min). Multiple sessions could occur per day (as many as they choose to participate in).
11) If a bird stops participating, the experimenter can give them yellow tube habituation trials to increase their motivation to participate in the actual experiment.

Criterion to pass: at least 17 of the most recent 20 trials correct with at least 8/10 or 9/10 correct in the most recent 2 sets of 10. Criterion is evaluated every trial such that an individual could pass in 20, 21, 32, etc trials.

Reversal (they only get 1 reversal)

- Always place the food in the previously unrewarded option
- Same methods as for the Initial Discrimination

INNOVATION: multiaccess log (experimental design after McCune et al., 2019b)

Apparatus: A wooden multiaccess box with 4 loci, one on top, front, and left and right sides. Each locus is covered by a clear plastic door that opens in a different way. The doors are labeled as: "A" on top of log, "B" on left side of log, "C" on front of log, and "D" on right side of log (counterclockwise if looking at log with chain at top).

Habituation

Enter data in data_xpop > tab: data_mabhabituation

Video record sessions when trying to get the bird to pass habituation criterion

Video file naming convention: A031-Y 2018-12-26 MABlog Habituation S7 T4

- Each bird receives the MAB in their aviary overnight with the doors fixed in the "open" position using rubber bands and maintenance diet food placed inside the open cavities. EXCEPTIONS: the following birds were not given the MAB in the aviary overnight, but on the same day before the MAB habituation trials started: Adobo, Yuca, Taquito, Xango. The following birds were not given the MAB before habituation trials, but rather after habituation trials started: Marisco, Cuervo, and Verbena. Door D had accidentally fallen shut on Kau's second day and Galandra's fourth day with MAB habituation. It was relocked open, however we are not sure whether they tried to open the door during this time, in which case they would have undocumented experience with opening this door, therefore we must omit door D from the analyses for these birds.
- The next day, put the wooden MAB in the aviary with a piece of goldfish (or other preferred food) in each compartment, DOORS LOCKED OPEN.
- Once the bird eats comfortably from ALL loci, attempt to get them to pass habituation criterion by recording whether the bird approaches within 3 minutes and eats comfortably from any locus on 2 consecutive trials (a trial is considered to restart after rebaiting the loci). Then they are ready to start testing. Rebait log between trials when bird is done eating/drinking water. If they eat from one locus and continue onto another immediately, don't disrupt them (flushing can create an association between the MAB and you flushing them instead of them receiving a reward for interacting with it). However, criterion must be met by conducting 2 consecutive trials where the bird obtains food after you've reset the wooden MAB with a food reward in each locus. You can rebait from within the aviary by blocking the bird's view with your body so that they can't see the apparatus being manipulated.
  - If the bird does not approach within 3 minutes, take the log out of the aviary and try again in a new session after a break, or the next day.
- How to score the "Ate food within 3 min" column in the data sheet:
  - 0 = did not eat the food from inside a locus within 3 min of the trial start time (came to the ground near the log but did not eat from a locus, or ate from the locus but it took longer than 3 minutes.)
  - 1 = ate the food from inside a locus within 3 min of the trial start time
  - -1 = did not participate (did not eat food inside a locus or touch a locus)
  - NOTE: when scoring an individual session (i.e., 1 session = 1 row in the data sheet) rather than an individual trial (because the bird is not yet participatory enough for trial level data), score each session according to the highest number they achieved across the whole session. For example, if there were 5 trials in the session and the bird ate from a locus within 3 min of the session start time, then score the whole session as 1.
- Criterion for ending habituation: a bird must obtain the food within 3 min on 2 consecutive trials.

Test preparation

ALWAYS PUT MAB ON GROUND SO THE CAMERA CAN VIEW ALL OPTIONS BETTER

Summary: Set-up wooden MAB out of sight of the bird, with a half piece of goldfish (so that they can be seen through the doors) in every compartment. Make sure the cracker in the front compartment (the drawer) is pushed to the front so the bird sees it clearly. Make sure you only put maintenance diet or small cracker pieces in the right compartment (the push door, locus "D") so the grackle can get them out under from the door when it pushes the door in. Place the log in the center of the aviary (and move all other objects away from it) so the front compartment (the drawer, locus "C") is facing toward the aviary door (so the camera at the front of the aviary can clearly see interactions with all options).

Testing

Enter data in data_xpop > tab: data_mab

Video record all sessions

Video file naming convention: A031-Y 2018-12-26 MABlog S7 T4

- Session = maximum 10 trials. A trial ends when the food is obtained or 15 min has elapsed, whichever comes first. If the latter, the next session is conducted after a break or on the following day.
- Initially, all 4 doors are closed and all compartments contain a piece of goldfish. A correct response is scored if the food is obtained, and the door from which it is obtained is noted.
  - If the bird does not come down to contact the box after 5 minutes of trial time, bait the ground with a small piece of food approximately 6 inches away from the box to encourage them to participate.
  - If the bird is on the ground when the 10 minute trial time ends, give the bird another 5 minutes to go to the box. Do not interrupt the bird if it is at the box when the trial time ends, wait for it to finish interacting and move to the perch or to its water dish on the ground - Note how long the trial was: 10 or 15 min (i.e., how long the individuals had the opportunity to learn about the apparatus).
- How to score the “correct choice” column in the data sheet:
  - 1 = used one of the loci to obtain the food (regardless of whether they actually ate the food and regardless of whether they touch [but don’t solve] other loci earlier in the trial).
  - 0 = the bird touched the box and/or loci, but doing so did not result in successfully opening a door (in this case, the session would time out and the log would be removed)
  - -1 = the bird did not touch the box or loci during the whole session. This trial is incomplete and is re-conducted until the bird scores a 1 or a 0.
- Criteria for solving one method: successfully obtain the food 3 times from a compartment. Once criterion is reached for one locus, lock that door open and empty it of food to make it non-functional.
- Criteria for ending the experiment:
  - When all 4 loci are non-functional,
  - if bird does not come to the ground within 10 min in 3 consecutive sessions when it is known that the bird is not afraid of the apparatus or experimenter (e.g., indicated by previous participation in this experiment) and when the sessions were not disrupted by external noise (note: sometimes a bird wasn’t participating because they were hesitant to approach the apparatus [in these cases, we continued with habituation to the pieces of the apparatus] or because they needed to re-habituate to the experimenter after catching for health checks),
  - or if bird does not obtain food within 10 min (or 15 min if the bird was on the ground at 10 mins) in 3 consecutive sessions (not including bait if food was put on the floor of the aviary to entice the bird to participate) when it is known that the bird is not afraid of the apparatus or experimenter and when the sessions were not disrupted by external noise.

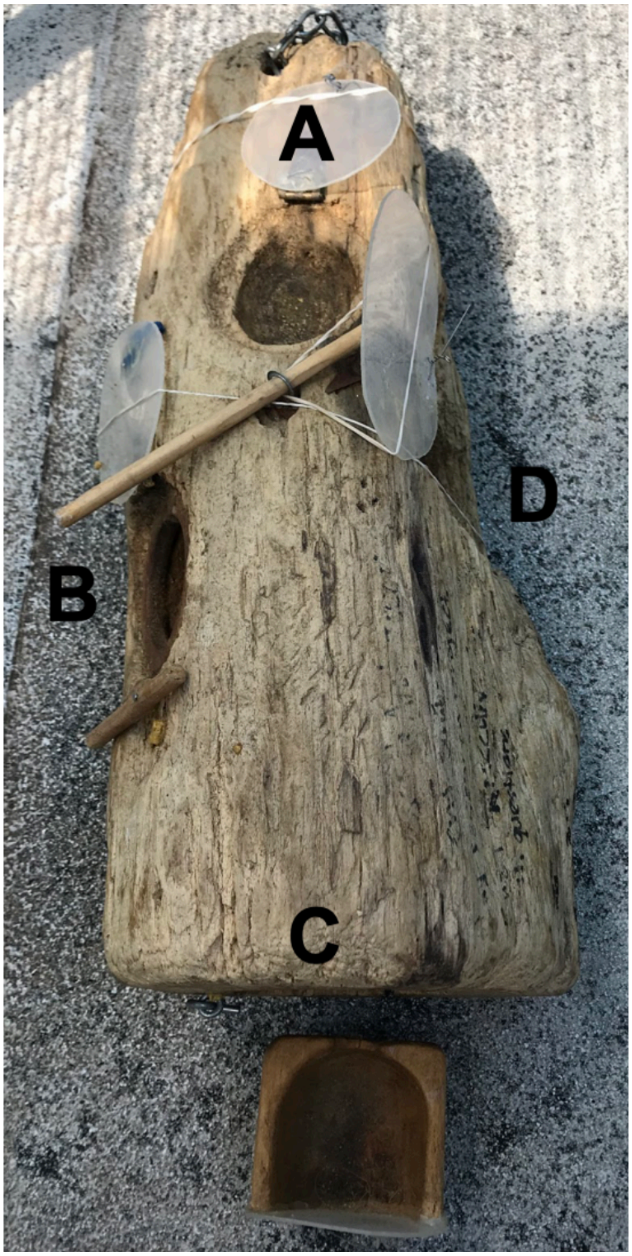

Figure SM5.1. For habituation, use rubber bands to secure A, B, and D doors open. C door can be pulled out and set on the ground. Then fill all with food.

Figure SM5.2. The doors are labeled as: “A” on top of log, “B” on left side of log, “C” on front of log, and “D” on right side of log.

Figure SM5.3. View of C door on the front of the log, showing the placement of the cracker right up at the front of the drawer so grackles can see it during trials.

EXPLORATION: environment (experimental design as in Mettke-Hofmann et al., 2009)

Summary: Time 1 occurs on a grackle's 8th day in the aviary or shortly thereafter (timing can be delayed due to not being able to run assays concurrently on several birds at a time if their aviary entry dates are close together). The bird's regular food is moved to one end of the aviary, away from the familiar/novel environments, and a motivation test begins the session. The bird is then exposed to first a familiar environment (45 min) and then a novel environment (45 min).

- All exploration assays are video recorded and take place with the experimenter out of view (at least 2 aviaries away)
- Enter data in data_xpop > tab: data_explore
- Video file naming convention:
  - A031-Y 2018-12-26 ExpEnv nov T1
  - A031-Y 2018-12-26 ExpEnv fam T2

Apparatus: the novel environment that will be placed inside the familiar environment (the aviary) is a tent (109cm wide by 58cm long by 46cm high; The Cat House https://nalaandcompany.com) with a zip open door that stays open using velcro.

Motivation test (not video recorded - move food to one end of the aviary): Place a piece of goldfish (or their most preferred food if not goldfish) in the center of the floor of the aviary (where the novel environment will be) and stay out of view of the bird for 5 min. If the grackle comes to eat the goldfish within 5 minutes, they are motivated to participate in the task and you can begin the session. If they do not come to take the cracker, wait 1 hour and try again. Scoring: 1 = the bird ate the food, 0 = the bird did not eat the food (enter data in the "CameToGroundForFoodBeforeTrial" column).

Time 1 Record 1 session per bird per environment (familiar first, then novel). Always record the familiar environment first.

1. Familiar environment

1) Move the regular food to the end of the aviary (against the back wall or door at the front), so they can still eat maintenance diet if they wish. Make sure there is no food near the tent area (even though there is no tent in this condition). Sweep up any maintenance diet that has been spilled in the area where the novel environment will be. Move all objects on the ground outside of the area delineated by the red stakes for the tent (see Figure SM5.5).
2) Conduct the motivation test (above).
3) Place a video camera outside of the aviary so that it views the entire floor. For the best view to estimate distance of the bird from the novel environment, make sure two of the tripod legs are against the back wall of the aviary aisle. The higher the camera is, the better the estimate of distance.
4) On a clean white board write:
    - The date
    - ID: X###XX, NAME (e.g., A046NG, Avocada)
    - Explore Environment
    - Time: 1 (or 2)
    - Condition: Familiar
    - Trial: X (X = how many times this scenario has been attempted for the individual)
    - Experimenter: XX (replace XX with the initials of the experimenter, e.g., CL)
5) Check that the camera battery has at least 45 minutes left. Start the camera, holding the white board in view in front of the camera for ~5 seconds, and set a timer for 45 minutes, then move out of view (at least 2 aviaries away) of the bird in this aviary for the whole trial time. At the end of the familiar trial, review the video to see if the grackle came to the floor. If the grackle did not come to the floor, it receives a ceiling value of 46 minutes in latency and the familiar trial should be attempted again (can occur immediately) until it is complete BEFORE conducting the novel trial.

2. Novel environment

1) Should occur immediately after the familiar environment trial (when the familiar trial was successful meaning that the bird came to the ground). The regular food remains at the end of the aviary away from where the tent will be so they can still eat maintenance diet if they wish. Make sure there is no food where the tent will be.
2) Conduct the motivation test (above).
3) Refresh the notes on the white board to reflect the new condition (i.e., Novel) and any other details that have changed.
4) Place the camera outside the aviary such that the inside of the tent and the rest of the aviary floor are visible. Check that the camera battery has at least 45 minutes left. Start the camera, holding the white board in view in front of the camera for ~5 seconds.
5) Place the tent (with its door velcro-ed open, leaving as big of an opening for the bird to walk through as possible) on the floor in the center of the aviary parallel to the door such that the tent door faces the aviary door and camera and make sure it is centered between the 4 red stakes in the ground (see Figure SM5.5). Ensure the tent is equidistant from the stakes in the ground that mark 20cm from its edges. Place one large aviary rock on each side of the tent so that its floor is flat (it is becoming permanently bent so the floor only partially touches the ground).
6) Set a timer for 45 minutes, remain out of view of the bird in this aviary for the whole trial time. The behaviors in this video will be entered into the data sheet when the videos are coded at a later date. If the grackle did not come to the floor, it receives a ceiling value of 46 minutes in latency.

Enter an event for Time 2 one week after Time 1 using the gtgrackles google calendar.

Time 2 (1 week after Time 1)

Repeat exactly as in Time 1.

Exceptions

- The tent door was closed for Xango's, Flan's, Alegria's and Chocolate's novel environment T2 assays so they would not have been likely to push the door aside to enter the tent.
- Experimenter came within two aviaries during Chocolate's T2 novel environment assay to test another bird.
- Motivation test was not conducted before novel environment due to experimenter error for Xango novel environment T1

Figure SM5.4. Novel environment (tent) set up in the aviary. As the edges have started to curl up with age, one side is lodged under the wall of the aviary and the other held down by the aviary rock.

Red stakes
mark 20 cm at 4 points for the object (connected by white line)

and 4 points for the tent (along blue line)

Figure SM5.5. Video coders mark lines on the Exploration videos to show that anything inside the white lines is within 20 cm of the object (familiar and novel) and that anything between the blue lines is within 20 cm of the tent (familiar and novel). For all conditions, ensure the water dishes are outside of the area of the blue lines (at the front or back of the aviary). For environment familiar, place the rocks in the position they will be in for the novel condition.

EXPLORATION: novel object (experimental design as in Mettke-Hofmann et al., 2009) - DISCONTINUED as of 9 Feb 2021

NOTE: this assay was discontinued as of 9 Feb 2021 because the Arizona exploration results came in. The results showed that only Exploration Environment is repeatable and that Exploration Object is not. Therefore, we will proceed only with the assay that is repeatable.

Summary: Time 1 occurs on a grackle's 8th day in the aviary or shortly thereafter (timing can be delayed due to not being able to run assays concurrently on several birds at a time if their aviary entry dates are close together). The bird is exposed to first a familiar object (45 min) and then a novel object (45 min) all occurring in the presence of their regular food (which is not next to the familiar/novel objects).

- All exploration assays are video recorded and take place with the experimenter out of view (at least 2 aviaries away)
- Enter data in data_xpop > tab: data_explore
- Video file naming convention:
    - A031-Y 2018-12-26 ExpObj nov T1
    - A031-Y 2018-12-26 ExpObj fam T2

Apparatus:

- Novel object: pink fuzzy wire bent into a random shape
- Familiar object: empty water dish

Motivation test = Place a piece of goldfish on the floor of the aviary. If the grackle comes to eat the goldfish within 5 minutes, they are motivated to participate in the task and you can begin the session. If they do not come to take the cracker, wait 1 hour and try again.

Time 1

Record 1 session per bird per object type (familiar first, then novel). Always record the familiar object session first.

Familiar object condition

1) Move the maintenance diet to one end of the aviary, away from where the object will be so they can still eat if they wish. Make sure there is no food in the area that the object will be in. Place a video camera outside of the aviary so that it views the entire floor. Move all objects on the ground outside of the area delineated by the red stakes for the object.
2) Conduct the motivation test.
3) If they pass the motivation test, start the camera (FOR THE EXPLORATION OBJECT EXPERIMENT ONLY: and place the familiar object on the floor in the center of the aviary. Ensure the object is equidistant from the stakes in the ground that mark 20cm from its edges.)
4) Set a timer for 45 minutes, remain out of view of the bird in this aviary for that whole trial time. At the end of the familiar trial, review the video to see if the grackle came to the floor. If the grackle did not come to the floor, it receives a ceiling value of 46 minutes in latency and the familiar trial should be attempted again (can occur immediately) until it is complete BEFORE conducting the novel trial.

Figures SM5.6 and SM5.7. Familiar object (empty water dish) (left) or novel object (right) placed in center of aviary for exploration test.

Novel object condition

1) Should occur immediately after the familiar object, on the same day.

2) Move the maintenance diet food to one end of the aviary away from where the object will be so they can still eat if they wish. Make sure there is no food in the area where the object will be.

3) Conduct the motivation test.
4) Place the novel object on the floor in the center of the aviary and make sure it is centered between the 4 red stakes in the ground (see Figure SM5.5). Ensure the object is equidistant from the stakes in the ground that mark 20cm from its edges. Place a video camera outside of the aviary so that it views the entire floor.
5) Start the camera and set a timer for 45 minutes, remain out of view of the bird in this aviary for that whole trial time. At the end of the trial, review the camera to see if the grackle came to the floor. If the grackle did not come to the floor, it receives a ceiling value of 46 minutes in latency.

Enter an event for Time 2 one week after Time 1 using the gtgrackles team google calendar.

Time 2 (1 week after Time 1)

Repeat exactly as in Time 1.

Exceptions: Experimenter came within two aviaries during Xango's T2 novel object assay to remove food from other aviaries.

REFERENCES

Audet J-N, Lefebvre L (2017) What's flexible in behavioral flexibility? *Behavioral Ecology*, **28**, 943–947. https://doi.org/10.1093/beheco/arx007

Auersperg AMI, Bayern AMP von, Gajdon GK, Huber L, Kacelnik A (2011) Flexibility in problem solving and tool use of kea and New Caledonian crows in a multi access box paradigm. *PLOS ONE*, **6**, e20231. https://doi.org/10.1371/journal.pone.0020231

Auersperg A, Szabo B, Von Bayern AM, Kacelnik A (2012) Spontaneous innovation in tool manufacture and use in a goffin's cockatoo. *Current Biology*, **22**, R903–R904.

Auguie B (2017) *gridExtra: Miscellaneous functions for "grid" graphics*.

Bates D, Mächler M, Bolker B, Walker S (2015) Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, **67**, 1–48. https://doi.org/10.18637/jss.v067.i01

Bates D, Maechler M, Bolker B (2012) lme4: Linear mixed-effects models using S4 classes (2011). R package version 0.999375-42.

Beran MJ (2015) The comparative science of "self-control": What are we talking about? *Frontiers in Psychology*, **6**. https://doi.org/10.3389/fpsyg.2015.00051

Bird CD, Emery NJ (2009) Insightful problem solving and creative tool modification by captive nontool-using rooks. *Proceedings of the National Academy of Sciences*, **106**, 10370–10375.

BirdLife_International (2018) Quiscalus mexicanus. *The IUCN Red List of Threatened Species 2018*, e.T22724308A132174807.

Blaisdell A, Seitz B, Rowney C, Folsom M, MacPherson M, Deffner D, Logan CJ (2021b) Do the more flexible individuals rely more on causal cognition? Observation versus intervention in causal inference in great-tailed grackles. *Peer Community Journal*, **1**.

Blaisdell A, Seitz B, Rowney C, Folsom M, MacPherson M, Deffner D, Logan CJ (2021a) 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 BM (2008) *Ecological models and data in r*. Princeton University Press.

Bond AB, Kamil AC, Balda RP (2007) Serial reversal learning and the evolution of behavioral flexibility in three species of north american corvids (gymnorhinus cyanocephalus, nucifraga columbiana, aphelocoma californica). *Journal of Comparative Psychology*, **121**, 372.

Carpenter B, Gelman A, Hoffman MD, Lee D, Goodrich B, Betancourt M, Brubaker M, Guo J, Li P, Riddell A (2017) Stan: A probabilistic programming language. *Journal of statistical software*, **76**.

Chapple DG, Simmonds SM, Wong BB (2012) Can behavioral and personality traits influence the success of unintentional species introductions? *Trends in ecology & evolution*, **27**, 57–64. https://doi.org/10.1016/j.tree.2011.09.010

Chow PKY, Lea SE, Leaver LA (2016) How practice makes perfect: The role of persistence, flexibility and learning in problem-solving efficiency. *Animal behaviour*, **112**, 273–283. https://doi.org/10.1016/j.anbehav.2015.11.014

Christensen AF (2000) The fifteenth-and twentieth-century colonization of the basin of mexico by the great-tailed grackle (quiscalus mexicanus). *Global Ecology and Biogeography*, **9**, 415–420.

Chuang A, Riechert SE (2021) Personality shifts are inconsistent across two rapid range expansions of a non-native spider. *Animal Behaviour*, **173**, 67–79. https://doi.org/10.1016/j.anbehav.2020.12.018

Cohen TM, Kumar RS, Nair M, Hauber ME, Dor R (2020) Innovation and decreased neophobia drive invasion success in a widespread avian invader. *Animal Behaviour*, **163**, 61–72. https://doi.org/10.1016/j.anbehav.2020.02.012

Collias EC, Collias NE (1964) The development of nest-building behavior in a weaverbird. *The Auk*, **81**, 42–52.

Eccard JA, Mazza V, Holland C, Stuart P (2022) The timid invasion: Behavioral adjustments favor range expansion in a non-native rodent. *bioRxiv*, 2022–06. https://doi.org/10.1101/2022.06.16.494985

Eddelbuettel D, François R (2011) Rcpp: Seamless R and C++ integration. *Journal of Statistical Software*, **40**, 1–18. https://doi.org/10.18637/jss.v040.i08

Fox RA, Ladage LD, Roth II TC, PRAVOSUDOV VV (2009) Behavioural profile predicts dominance status in mountain chickadees, Poecile gambeli. *Animal Behaviour*, **77**, 1441–1448. https://doi.org/10.1016/j.anbehav.2009.02.022

Gabry J, Češnovar R (2021) *Cmdstanr: R interface to 'CmdStan'.*

Gamer M, Lemon J, Gamer MM, Robinson A, Kendall's W (2012) Package "irr." *Various coefficients of interrater reliability and agreement.*

Griffin AS, Diquelou MC (2015) Innovative problem solving in birds: A cross-species comparison of two highly successful passerines. *Animal Behaviour*, **100**, 84–94.

Griffin AS, 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

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

Griffin AS, Netto K, Peneaux C (2017) Neophilia, innovation and learning in an urbanized world: A critical evaluation of mixed findings. *Current Opinion in Behavioral Sciences*, **16**, 15–22. https://doi.org/10.1016/j.cobeha.2017.01.004

Hadfield J (2010a) MCMC methods for multi-response generalized linear mixed models: The MCMCglmm r package. *Journal of Statistical Software*, **33**, 1–22. https://doi.org/10.18637/jss.v033.i02

Hadfield J (2010b) MCMCglmm: Markov chain monte carlo methods for generalised linear mixed models. *Tutorial for MCMCglmm package in R*, **125**.

Hadfield J (2014) MCMCglmm course notes.

Haemig PD (2011) Introduction of the great-tailed grackle by aztec emperor auitzotl: Four-stage analysis with new information. *Ardeola*, **58**, 387–397. https://doi.org/10.13157/arla.58.2.2011.387

Haemig PD (2012) Introduction of the great-tailed grackle (quiscalus mexicanus) by aztec emperor auitzotl: Provenance of the historical account. *The Auk*, **129**, 70–75. https://doi.org/10.1525/auk.2011.11058

Haemig P (2014) Aztec introduction of the great-tailed grackle in ancient mesoamerica: Formal defense of the sahaguntine historical account. *NeoBiota*, **22**, 59–75. https://doi.org/10.3897/neobiota.22.6791

Hartig F (2019) *DHARMa: Residual diagnostics for hierarchical (multi-level / mixed) regression models.*

Hutcheon JA, Chiolero A, Hanley JA (2010) Random measurement error and regression dilution bias. *Bmj*, **340**, c2289. https://doi.org/10.1136/bmj.c2289

Isaksson E, Utku Urhan A, Brodin A (2018) High level of self-control ability in a small passerine bird. *Behavioral ecology and sociobiology*, **72**, 1–7. https://doi.org/10.1007/s00265-018-2529-z

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

Knolle F, Goncalves RP, Davies EL, Duff AR, Morton AJ (2019) Response-inhibition during problem solving in sheep. *International Journal of Comparative Psychology*, **32**. https://doi.org/10.46867/ijcp.2019.32.00.15

Kotrschal A, Rogell B, Bundsen A, Svensson B, Zajitschek S, Brännström I, Immler S, Maklakov AA, Kolm N (2013) Artificial selection on relative brain size in the guppy reveals costs and benefits of evolving a larger brain. *Current Biology*, **23**, 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**, 187–196.

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

Laughlin RE, Grant TL, Williams RW, Jentsch JD (2011) Genetic dissection of behavioral flexibility: Reversal learning in mice. *Biological psychiatry*, **69**, 1109–1116.

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

Lefebvre L, Whittle P, Lascaris E, Finkelstein A (1997) Feeding innovations and forebrain size in birds. *Animal Behaviour*, **53**, 549–560. https://doi.org/10.1006/anbe.1996.0330

Lissek S, Diekamp B, Güntürkün O (2002) Impaired learning of a color reversal task after NMDA receptor blockade in the pigeon (columbia livia) associative forebrain (neostriatum caudolaterale). *Behavioral Neuroscience*, **116**, 523.

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

Logan CJ (2016b) Behavioral flexibility and problem solving in an invasive bird. *PeerJ*, **4**, e1975. https://doi.org/10.7717/peerj.1975

Logan CJ, Avin S, Boogert N, Buskell A, Cross FR, Currie A, Jelbert S, Lukas D, Mares R, Navarrete AF, Shigeno S, Montgomery SH (2017) Beyond Brain Size. *bioRxiv*, 145334. https://doi.org/10.1101/145334

Logan CJ, Lukas D, Blaisdell AP, Johnson-Ulrich Z, MacPherson M, Seitz B, Sevchik A, McCune KB (2023a) Behavioral flexibility is manipulable and it improves flexibility and innovativeness in a new context. *EcoEvoRxiv, version 5 of this preprint has been peer reviewed and recommended by Peer Community In Ecology [https://doi.org/10.24072/pci.ecology.100407].* https://doi.org/10.32942/osf.io/5z8xs

Logan C, Lukas D, Blaisdell A, Johnson-Ulrich Z, MacPherson M, Seitz B, Sevchik A, McCune K (2023b) Data: Is behavioral flexibility manipulatable and, if so, does it improve flexibility and problem solving in a new context? *Knowledge Network for Biocomplexity*, **Data package**. https://doi.org/10.5063/F1BR8QNC

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

Logan CJ, McCune KB, Chen N, Lukas D (2020) Implementing a rapid geographic range expansion - the role of behavior and habitat changes. *In principle acceptance by PCI Ecology of the version on 6 Oct 2020.*

Logan CJ, McCune K, MacPherson M, Johnson-Ulrich Z, Rowney C, Seitz B, Blaisdell A, Deffner D, Wascher C (2021) Are the more flexible great-tailed grackles also better at behavioral inhibition? *PsyArXiv*. https://doi.org/10.31234/osf.io/vpc39

Logan CJ, McCune KB, Marfori Z, Rolls C, Lukas D (2023c) Data: Implementing a rapid geographic range expansion - the role of behavior changes. *Knowledge Network for Biocomplexity*, **Data package**. https://doi.org/10.5063/F10C4T7T

Logan CJ, Shaw R, Lukas D, McCune KB (2022) How to succeed in human modified environments. *In principle acceptance by PCI Ecology of the version on 8 Sep 2022.*

Lukas D, McCune K, Blaisdell A, Johnson-Ulrich Z, MacPherson M, Seitz B, Sevchik A, Logan C (2022) Behavioral flexibility is manipulatable and it improves flexibility and problem solving in a new context: Post-hoc analyses of the components of behavioral flexibility. *EcoEvoRxiv*. https://doi.org/10.32942/osf.io/4ycps

MacLean EL, Hare B, Nunn CL, Addessi E, Amici F, Anderson RC, Aureli F, Baker JM, Bania AE, Barnard AM, Boogert NJ, Brannon EM, Bray EE, Bray J, Brent LJN, Burkart JM, Call J, Cantlon JF, Cheke LG, Clayton NS, Delgado MM, DiVincenti LJ, Fujita K, Herrmann E, Hiramatsu C, Jacobs LF, Jordan KE, Laude JR, Leimgruber KL, Messer EJE, Moura AC de A, Ostojić L, Picard A, Platt ML, Plotnik JM, Range F, Reader SM, Reddy RB, Sandel AA, Santos LR, Schumann K, Seed AM, Sewall KB, Shaw RC, Slocombe KE, Su Y, Takimoto A, Tan J, Tao R, Schaik CP van, Virányi Z, Visalberghi E, Wade JC, Watanabe A, Widness J, Young JK, Zentall TR, Zhao Y (2014) The evolution of self-control. *Proceedings of the National Academy of Sciences*, **111**, E2140–E2148. https://doi.org/10.1073/pnas.1323533111

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

McCune KB (2018) Cognition gone wild: A test of the social intelligence hypothesis in wild birds. PhD thesis Thesis.

McCune K, Blaisdell A, Johnson-Ulrich Z, Lukas D, MacPherson M, Seitz B, Sevchik A, Logan C (2022) Repeatability of performance within and across contexts measuring behavioral flexibility. *EcoEvoRxiv*. https://doi.org/10.32942/osf.io/kevqp

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

McCune KB, MacPherson M, Rowney C, Bergeron L, Folsom M, Logan C (2019b) 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.*

McElreath R (2020a) *Rethinking: Statistical rethinking book package.*

McElreath R (2020b) *Statistical rethinking: A bayesian course with examples in r and stan.* Chapman; Hall/CRC, Boca Raton, FL. https://doi.org/10.1201/9780429029608

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**, 482–492.

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

Mikhalevich I, Powell R, Logan C (2017) Is behavioural flexibility evidence of cognitive complexity? How evolution can inform comparative cognition. *Interface Focus*, **7**, 20160121. https://doi.org/10.1098/rsfs.2016.0121

Morand-Ferron J, Cole EF, Rawles JE, Quinn JL (2011) Who are the innovators? A field experiment with 2 passerine species. *Behavioral Ecology*, **22**, 1241–1248. https://doi.org/10.1093/beheco/arr120

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

R Core Team (2023) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Réale D, Reader SM, Sol D, McDougall PT, Dingemanse NJ (2007) Integrating animal temperament within ecology and evolution. *Biological reviews*, **82**, 291–318.

Revelle W (2017) *Psych: Procedures for psychological, psychometric, and personality research*. Northwestern University, Evanston, Illinois.

Rodrigo A, Ávila-Chauvet L, Buriticá J (2021) Daily patterns of foraging and aggressive behaviors in great-tailed grackle (quiscallus mexicanus) at an urban patch with availability or absence of resources. *bioRxiv*. https://doi.org/10.1101/2021.06.14.448443

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

Sarkar D (2008) *Lattice: Multivariate data visualization with r*. Springer, New York.

Seitz KAM Benjamin M. AND McCune (2021) Using touchscreen equipped operant chambers to study animal cognition. Benefits, limitations, and advice. *PLOS ONE*, **16**, 1–18. https://doi.org/10.1371/journal.pone.0246446

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

Sol D, Duncan RP, Blackburn TM, 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**, 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**, 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**, 763–769. https://doi.org/10.1098/rspb.2006.3765

Sol D, Timmermans S, Lefebvre L (2002) Behavioural flexibility and invasion success in birds. *Animal behaviour*, **63**, 495–502. https://doi.org/10.1006/anbe.2001.1953

Stan Development Team (2020) RStan: The R interface to Stan.

Stoffel MA, Nakagawa S, Schielzeth H (2017) rptR: Repeatability estimation and variance decomposition by generalized linear mixed-effects models. *Methods in Ecology and Evolution*, **8**, 1639–1644.

Storks L, Leal M (2020) Thinking outside the box: Problem-solving in free-living lizards. *Behavioral Ecology and Sociobiology*, **74**, 1–9. https://doi.org/10.1007/s00265-020-02852-x

Summers J, Lukas D, Logan C, Chen N (2023) The role of climate change and niche shifts in divergent range dynamics of a sister-species pair. *Peer Community Journal*. https://doi.org/10.24072/pcjournal.248

Szabo B, Damas-Moreira I, Whiting MJ (2020) Can cognitive ability give invasive species the means to succeed? A review of the evidence. *Frontiers in Ecology and Evolution*, **8**. https://doi.org/10.3389/fevo.2020.00187

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

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

Troisi CA, Cooke AC, Davidson GL, Hera I de la, Reichert MS, Quinn JL (2020) No evidence for cross-contextual consistency in spatial learning and behavioural flexibility in a passerine. *BioRxiv*, 2020–09. https://doi.org/10.26451/abc.08.03.08.2021

Vehtari A, Gelman A, Simpson D, Carpenter B, Bürkner P-C (2021) Rank-normalization, folding, and localization: An improved rhatfor assessing convergence of MCMC (with discussion). *Bayesian Analysis*.

Wehtje W (2003) The range expansion of the great-tailed grackle (quiscalus mexicanus gmelin) in north america since 1880. *Journal of Biogeography*, **30**, 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, François R, Henry L, Müller K (2021) *Dplyr: A grammar of data manipulation*.

Wickham H, Vaughan D, Girlich M (2023) *Tidyr: Tidy messy data*.

Wright TF, Eberhard JR, Hobson EA, Avery ML, Russello MA (2010) Behavioral flexibility and species invasions: The adaptive flexibility hypothesis. *Ethology Ecology & Evolution*, **22**, 393–404. https://doi.org/10.1080/03949370.2010.505580

Xie Y (2013) Knitr: A general-purpose package for dynamic report generation in r. *R package version*, **1**.

Xie Y (2017) *Dynamic documents with r and knitr*. Chapman; Hall/CRC. https://doi.org/10.1201/b15166

Xie Y (2018) Knitr: A comprehensive tool for reproducible research in r. In: *Implementing reproducible research*, pp. 3–31. Chapman; Hall/CRC. https://doi.org/10.1201/9781315373461-1

Xie Y (2023) *formatR: Format r code automatically*.

Zhu H (2021) *kableExtra: Construct complex table with 'kable' and pipe syntax*.