

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

3 Logan CJ^{1*} McCune KB² LeGrande-Rolls C¹ Marfori Z¹ Hubbard J³
4 Lukas D¹

5 2023-04-11



7
8 **Affiliations:** 1) Max Planck Institute for Evolutionary Anthropology, 2) University of California Santa
9 Barbara, 3) Animal Behavior Graduate Group, University of California Davis. *Corresponding author:
10 corina_logan@eva.mpg.de

11
12 **This is the post-study manuscript of the preregistration that was pre-study peer reviewed and
13 received an In Principle Recommendation on 16 Dec 2021 by:**

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



17
18 **Preregistration:** [html](#), [pdf](#), [rmd](#)
19 **Post-study manuscript** (submitted to PCI Ecology for post-study peer review on 12 April 2023): preprint
20 pdf at EcoEvoRxiv, [html](#), [rmd](#)

21 **ABSTRACT**

22 It is generally thought that behavioral flexibility, the ability to change behavior when circumstances change,
23 plays an important role in the ability of species to rapidly expand their geographic range. Great-tailed
24 grackles (*Quiscalus mexicanus*) are a social, polygamous species that is rapidly expanding its geographic
25 range by settling in new areas and habitats. They are behaviorally flexible and highly associated with
26 human-modified environments, eating a variety of human foods in addition to foraging on insects and on the
27 ground for other natural food items. They offer an opportunity to assess the role of behavior change over
28 the course of their expansion. We compare behavior in wild-caught grackles from two populations across
29 their range (an older population in the middle of the northern expansion front: Tempe, Arizona, and a more
30 recent population on the northern edge of the expansion front: Woodland, California) to investigate whether

31 certain behaviors (flexibility, innovativeness, exploration, and persistence) have higher averages and variances
32 in the newer or older population. We find that grackles in the edge population were more innovative and
33 less exploratory, and that there were no population differences in flexibility (measured by reversal learning)
34 or persistence (the proportion of trials participated in). Results elucidate that the rapid geographic range
35 expansion of great-tailed grackles is associated with individuals differentially expressing particular behaviors.
36 Our findings highlight the value of population studies and of breaking down cognitive concepts into direct
37 measures of individual abilities to better understand how species might adapt to novel circumstances.

38 INTRODUCTION

39 It is generally thought that behavioral flexibility, the ability to change behavior when circumstances change
40 through packaging information and making it available to other cognitive processes (see Mikhalevich et al.,
41 2017 for theoretical background on our flexibility definition), plays an important role in the ability of a
42 species to rapidly expand their geographic range (Chow et al., 2016; Griffin & Guez, 2014; e.g., Lefebvre et
43 al., 1997; Sol et al., 2002, 2005, 2007; Sol & Lefebvre, 2000). These ideas predict that flexibility, exploration,
44 and innovation (creating new behaviors or using existing behaviors in a new context, Griffin & Guez, 2014)
45 facilitate the expansion of individuals into completely new areas and that their role diminishes after a certain
46 number of generations (Wright et al., 2010). In support of this, experimental studies have shown that latent
47 abilities are primarily expressed in a time of need (A. Auersperg et al., 2012; Bird & Emery, 2009; Laumer
48 et al., 2018; Manrique & Call, 2011; e.g., Taylor et al., 2007). Therefore, we do not expect the founding
49 individuals who initially dispersed out of their original range to have unique behavioral characteristics that
50 are passed on to their offspring. Instead, we expect that the actual act of continuing a range expansion relies
51 on flexibility, exploration, innovation, and persistence, and that these behaviors are therefore expressed more
52 on the edge of the expansion range where there have not been many generations to accumulate relevant
53 knowledge about the environment.

54 To determine whether behavior is involved in a rapid geographic range expansion, direct measures of individual
55 behavioral abilities must be collected in populations across the range of the species (see the discussion on
56 the danger of proxies of flexibility in Logan et al., 2018). Our study aims to test whether behavioral flexibility
57 plays a role in the rapid geographic range expansion of great-tailed grackles (*Quiscalus mexicanus*). Great-
58 tailed grackles are behaviorally flexible (Logan, 2016a), rapidly expanding their geographic range (Wehtje,
59 2003), and highly associated with human-modified environments (Johnson & Peer, 2001), thus offering an
60 opportunity to assess the role of behavior changes over the course of their expansion. This social, polygamous
61 species eats a variety of human foods in addition to foraging on insects and on the ground for other natural
62 food items (Johnson & Peer, 2001). This feature increases the ecological relevance of comparative cognition
63 experiments that measure individual behavior abilities: grackles eat at outdoor cafes, from garbage cans, and
64 on crops. As such, they generally gain experience in the wild with approaching and opening novel objects
65 to seek food (e.g., attempting to open a ketchup packet at an outdoor cafe, climbing into garbage cans to
66 get french fries at the zoo, dunking sugar packets in water), which makes the tests involving human-made
67 apparatuses ecologically relevant for this species.

68 We aim to compare behavior in wild-caught great-tailed grackles from two populations across their range
69 (an older population in the middle of the northern expansion front: Tempe, Arizona using previously published
70 data from Logan et al. (2023), and a more recent population on the northern edge of the expansion front:
71 Woodland, California) (Figure 1, Table 1). We investigate whether certain behaviors have higher
72 averages and variances in the edge population relative to the older population. Specifically, we investigate
73 behavioral flexibility measured as reversal learning of food-filled colored tube preferences (Logan, 2016b;
74 Logan et al., 2023); innovativeness, measured as the number of loci they solve to access food on a puzzlebox
75 (AMI. Auersperg et al., 2011; Logan et al., 2023); exploration, measured as the latency to approach a novel
76 environment in the absence of nearby food (McCune KB et al., 2019; Mettke-Hofmann et al., 2009); and per-
77 sistence, measured as the proportion of sessions they participate in during the flexibility and innovativeness
78 experiments (Figure 2).

79 There could be multiple mechanisms underpinning the results, however our aim is to narrow down the role of
80 changes in behavior in the range expansion of great-tailed grackles. Our results demonstrate that the rapid

81 geographic range expansion of great-tailed grackles is associated with individuals differentially expressing
 82 particular behaviors in the edge compared to the older population.

83 **Table 1.** Population characteristics for the field sites. The number of generations at a site is based on a
 84 generation length of 5.6 years for this species [BirdLife International (2018); note that this species starts
 85 breeding at age 1] and on the first year in which this species was reported (or estimated) to breed at each
 86 location (Woodland, California: Yolo Audubon Society's newsletter *The Burrowing Owl* from July 2004; and
 87 Tempe, Arizona: estimated based on 1945 first-sighting report in nearby Phoenix, Arizona (Wehtje, 2004)
 88 to which we added 6 years to account for the average time between first-sighting and first-breeding - see
 89 Table 3 in Wehtje (2003). The average number of generations was calculated up to 2020, the final year of
 90 data collection in Tempe, and 2022, the final year of data collection in Woodland.

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

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

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

112 METHODS

113 Sample

114 Great-tailed grackles are caught in the wild in Woodland and in the Bufferlands of Sacramento, California.
 115 Some of our banded individuals were found at both sites, therefore we consider this one population. We aim
 116 to bring adult grackles, rather than juveniles, temporarily into the aviaries for behavioral choice tests to avoid
 117 the potential confound of variation in cognitive development due to age, as well as potential variation in fine
 118 motor-skill development (e.g., holding/grasping objects; early-life experience plays a role in the development
 119 of both of these behaviors; e.g., Collias & Collias (1964), Rutz et al. (2016)) with variation in our target

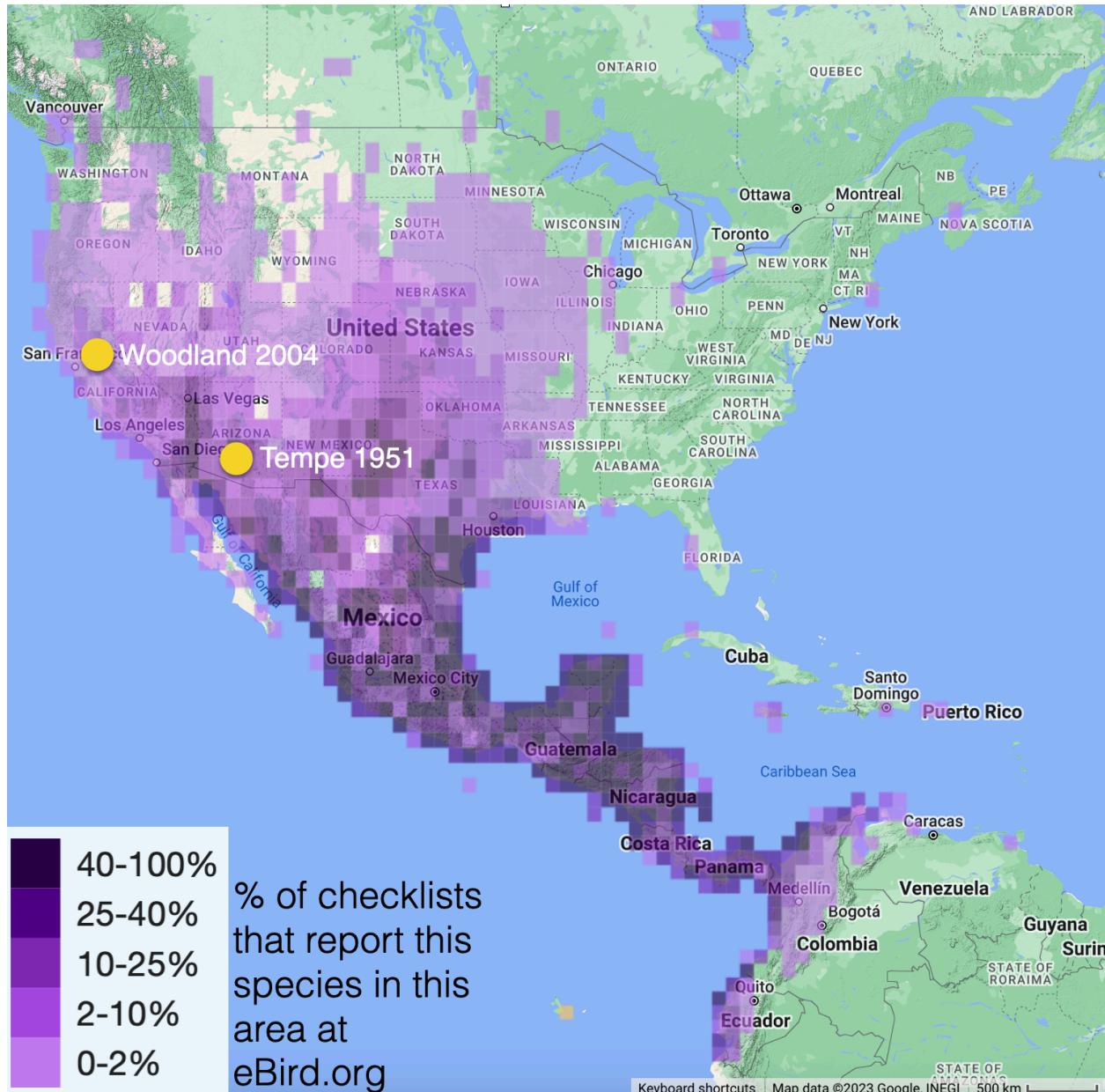


Figure 1: Great-tailed grackle field sites: Woodland is a recently established population (first breeding at the trapping location recorded in 2004) on the northern edge of the range, and Tempe is an older population (established in 1951) in the middle of the northern expansion front. Data from eBird.org).

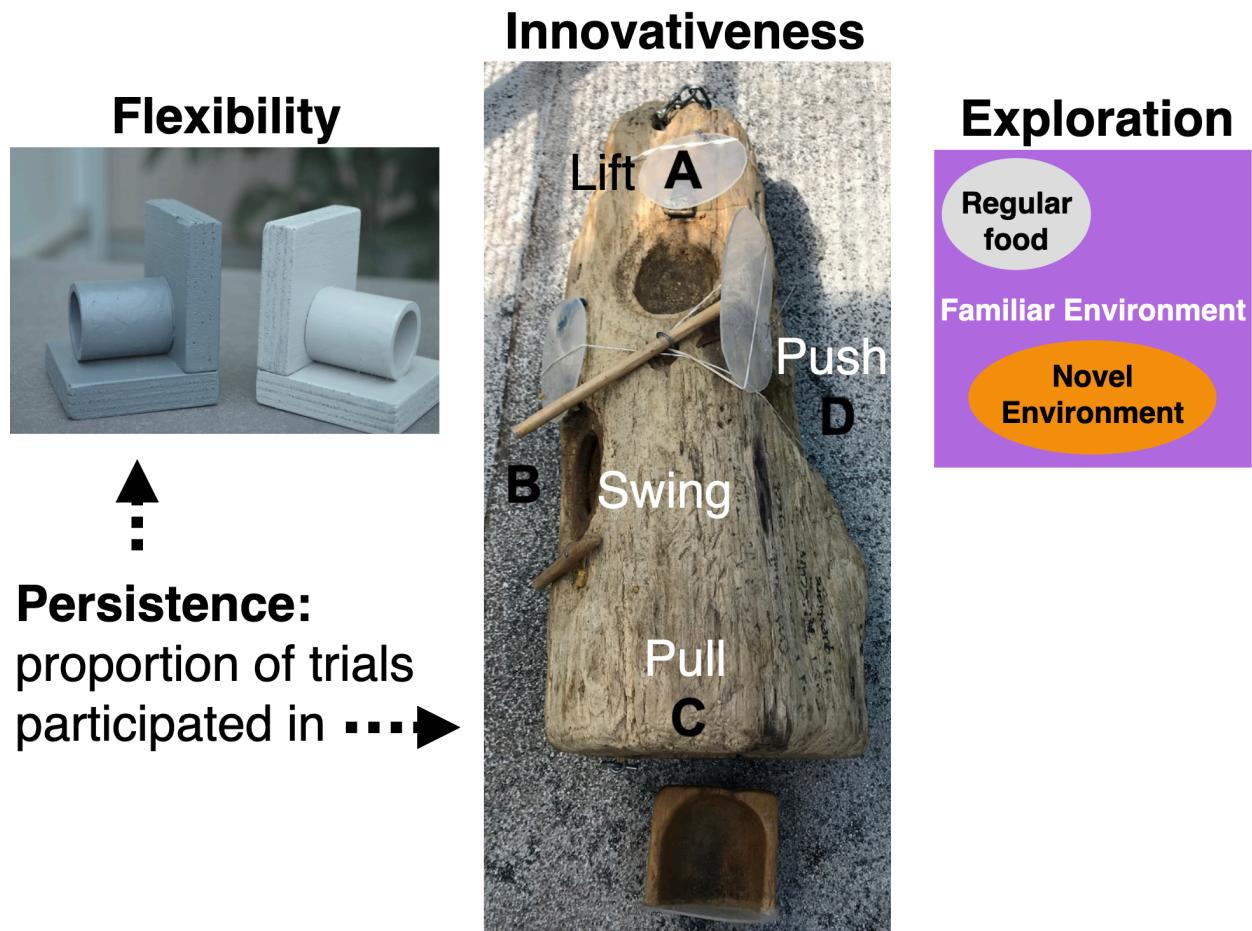


Figure 2: Experimental protocol. Great-tailed grackles from the older and newer populations are tested for their: (top left) flexibility (number of trials to reverse a previously learned color tube-food association); (middle) innovativeness (number of options [lift, swing, pull, push] solved to obtain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

120 variables of interest. However, due to difficulties in trapping this species at this site, we also tested some
 121 juveniles. This should not pose a problem because we found that the two juveniles (Taco and Chilaquile)
 122 we tested in the Tempe population did not perform differently from adults (Blaisdell et al., 2021; Logan et
 123 al., 2021; Logan et al., 2023; Seitz, 2021). Adults are identified from their eye color, which changes from
 124 brown to yellow upon reaching adulthood (Johnson & Peer, 2001). We apply colored leg bands in unique
 125 combinations for individual identification. Some individuals (33) are brought temporarily into aviaries for
 126 behavioral choice tests, and then are released back to the wild at their point of capture. We catch grackles
 127 with walk-in traps and mist nets. Mist nets decrease the likelihood of a selection bias for exploratory and
 128 bold individuals because grackles cannot see the trap. Grackles are individually housed in an aviary (each
 129 244 cm long by 122 cm wide by 213 cm tall) for three weeks to six months where they have *ad lib* access to
 130 water at all times and are fed Mazuri Small Bird maintenance diet *ad lib* during non-testing hours (minimum
 131 20 h per day), and various other food items (e.g., peanuts, bread, goldfish crackers) during testing (up to
 132 4 h per day per bird). Individuals Are given three to four days to habituate to the aviaries and then their
 133 test battery begins on the fourth or fifth day (birds were usually tested six days per week, therefore if their
 134 fourth day occurs on a day off, they are tested on the fifth day instead).

135 While our ideal plan was to conduct the same tests at an additional field site in Central America, due to
 136 restrictions around COVID-19 and also to issues with sexual abuse at the planned field site, it was not
 137 possible for us to accomplish this goal within our current funding period.

138 We test as many great-tailed grackles as we can during the 2 years we spend at each of our field sites given
 139 that the birds are only brought into the aviaries during the non-breeding season (September through April).
 140 It is time intensive to conduct the aviary test battery (3 weeks-6 months per bird), therefore we aim to
 141 meet the minimum sample sizes in Supplementary Material Table SM1. We aim for an equal sex ratio of
 142 subjects (50% female) and achieved an overall 47% female (this percentage differs depending on the test).
 143 We expected to test 20 grackles per site. See the `gxpbehaviorhabitat_data_testhistory.csv` data sheet at
 144 Logan CJ et al. (2023) for a list of the order of experiments for each individual at the Woodland site, and
 145 `g_flexmanip_data_AllGrackleExpOrder.csv` at C. Logan et al. (2023) for the Tempe grackles.

146 Data collection stopping rule

147 We stop collecting data on wild-caught great-tailed grackles once we complete one year at the site and meet
 148 our minimum sample size.

149 Protocols and open materials

- 150 • Experimental protocols are online [here](#).
- 151 • **Flexibility** protocol (from Logan et al., 2023) using reversal learning with color tubes. Grackles are
 152 first habituated to a yellow tube and trained to search for hidden food. A light gray tube and a dark
 153 gray tube are placed on the table or floor: one color always contains a food reward (not visible by the
 154 bird) while the other color never contains a reward. The bird is allowed to choose one tube per trial.
 155 An individual is considered to have a preference if it chose the rewarded option at least 85% of the time
 156 (17/20 correct) in the most recent 20 trials (with a minimum of 8 or 9 correct choices out of 10 on the
 157 two most recent sets of 10 trials). We use a sliding window in 1-trial increments to calculate whether
 158 they passed after their first 20 trials. Once a bird learns to prefer one color, the contingency is reversed:
 159 food is always in the other color and never in the previously rewarded color. The flexibility measure is
 160 how many trials it takes them to reverse their color preference using the same passing criterion.
- 161 • **Innovativeness** protocol AMI. Auersperg et al. (2011) using a multiaccess log. Grackles are first
 162 habituated to the log apparatus with all of the doors locked open and food inside each locus. After
 163 habituation, the log, which has four ways of accessing food (pull drawer, push door, lift door up, swing
 164 door out), is placed on the ground and grackles are allowed to attempt to solve or successfully solve one
 165 option per trial. Once a bird has successfully solved an option three times, it becomes non-functional

166 (the door is locked open and there is no food at that locus). The experiment ends when all four loci
167 become non-functional, if a bird does not come to the ground within 10 min in three consecutive test
168 sessions, or if a bird does not obtain the food within 10 min (or 15 min if the bird was on the ground
169 at 10 min) in three consecutive test sessions.

170 • **Persistence** is measured as the proportion of trials participated in during the flexibility and innova-
171 tiveness experiments (after habituation, thus it is not confounded with neophobia). The higher the
172 number, the more persistent they are. This measure indicates that those birds who do not participate
173 as often are less persistent in terms of their persistence with engaging with the task. We generally offer
174 a grackle the chance to participate in a trial for 5 min. If they do not participate within that time, we
175 record -1 in the data sheet, the apparatus is removed and the trial is re-attempted later.

176 • **Exploration** is measured as the latency to approach within 20 cm of a novel environment inside of
177 their familiar aviary environment, averaged across Time 1 (on the individual's 8th day in the aviary)
178 and Time 2 (1 week after Time 1). The bird's regular food is moved to one end of the aviary, away
179 from the novel environment, and a motivation test precedes the session. The bird is then exposed to
180 first a familiar environment (45 min) and then a novel environment (45 min). If an individual does not
181 approach within 20 cm, it is given a latency of 2701 sec (45 min plus 1 sec).

182 Open data

183 The data and code are publicly available at the Knowledge Network for Biocomplexity's data repository
184 (Logan CJ et al., 2023).

185 Randomization and counterbalancing

186 **Experimental order:** The order of experiments, reversal learning or multiaccess log, is counterbalanced
187 across birds.

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

196 Analyses

197 We use **simulations** and design customized **models** to determine what sample sizes allow us to detect
198 differences between sites (Supplementary Material 2; see chapter 5.3 in Bolker (2008) for why simulations
199 perform more powerful power analyses). We do not **exclude** any data, and data that are **missing** (e.g. if a
200 bird participates in one of the two experiments) for an individual in a given experiment, then this individual
201 is not included in that analysis. Analyses are conducted in R [current version 4.1.2; R Core Team (2017)] and
202 Stan (version 2.18, Carpenter et al., 2017) using the following packages: psych (Revelle, 2017), irr (Gamer
203 et al., 2012), rthinking (McElreath, 2020), rstan (Stan Development Team, 2020), knitr (Xie, 2013, 2017,
204 2018), dplyr (Wickham et al., 2021), tidyR (Wickham et al., 2023), cmdstanr (Gabry & Češnovar, 2021),
205 DHARMA (Hartig, 2019), lme4 (Bates et al., 2012; Bates et al., 2015), and Repp (Eddelbuettel & François,
206 2011). Interobserver reliability scores indicate high agreement across coders for all dependent variables (see
207 Supplementary Material 3 for details).

208 **Flexibility analyses**

209 **Model and simulation**

210 We modify the reversal learning Bayesian model in Blaisdell et al. (2021) to simulate and analyze population
 211 differences in reversal learning, and calculate our ability to detect differences between populations. The model
 212 accounts for every choice made in the reversal learning experiment and updates the probability of choosing
 213 either option after the choice is made depending on whether that choice contains a food reward or not. It
 214 does this by updating three main components for each choice: an attraction score, a learning rate (ϕ), and
 215 a rate of deviating from learned attractions (λ).

216 As in Blaisdell et al. (2021), we, too, use previously published data on reversal learning of color tube prefer-
 217 ences in great-tailed grackles in Santa Barbara, California (Logan, 2016b) to inform the model modifications.
 218 We modified the Blaisdell et al. (2021) model in a two ways: 1) we set the initial attraction score assigned
 219 to option 1 and option 2 (the light gray and dark gray tubes) to 0.1 rather than 0.0. This change assumes
 220 that there would be some inclination (rather than no inclination) for the bird to approach the tubes when
 221 they are first presented because they are previously trained to expect food in tubes. This also allows the
 222 attraction score to decrease when a non-rewarded choice is made near the beginning of the experiment. With
 223 the previous initial attraction scores set to zero, a bird would be expected to choose the rewarded option in
 224 100% of the trials after the first time it chose that option (attraction cannot be lower than zero, and choice
 225 is shaped by the ratio of the two attractions so that when one option is zero and the other is larger than
 226 zero, the ratio will be 100% for the rewarded option). 2) We changed the updating so that an individual
 227 only changes the attraction toward the option they chose in that trial (either decreasing their attraction
 228 toward the unrewarded option or increasing their attraction toward the rewarded option). Previously, both
 229 attractions were updated after every trial, assuming that individuals understand that the experiment is set
 230 up such that one option is always rewarded. For our birds, we instead assumed that individuals will focus on
 231 their direct experience rather than making abstract assumptions about the test. Our modification resulted
 232 in needing a higher ϕ to have the same learning rate as a model where both attraction scores update after
 233 every trial. This change also appears to better reflect the performance of the Santa Barbara grackles, because
 234 they had higher ϕ values, which, in turn, meant lower λ values to reflect the performance during their initial
 235 learning. These lower λ values better reflect the birds' behavior during the first reversal trials: a large λ
 236 value means that birds continue to choose the now unrewarded option almost 100% of the time, whereas the
 237 lower λ values mean that birds start to explore the rewarded option relatively soon after the switch of the
 238 rewarded option.

239 We first reanalyze the Santa Barbara grackle data to obtain the phi and lambda values with this revised
 240 model, which informs our expectations of what a site's mean and variance might be. Then we run simulations,
 241 where we determined that we wanted to make the previously mentioned modifications to the stan (Team et
 242 al., 2019) model [in R, current version 4.1.2; R Core Team (2017)]. This model is used to analyze the actual
 243 data after it is collected.

244 **Innovation analysis**

245 **Model and simulation**

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

250 $\text{locisolved} \sim \text{Binomial}(4, p)$ *[likelihood]*

251 $\text{logit}(p) \sim \alpha[\text{site}]$ *[model]*

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

256 $\alpha \sim \text{Normal}(0,1)$ *[a prior]*

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

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

266 Exploration analysis

267 Model and simulation

268 We model the average latency to approach a novel environment and compare these between sites. We
 269 simulate data and set the model as follows:

270 $\text{latency} \sim \text{gamma-Poisson}(\lambda_i, \phi)$ [*likelihood*]

271 $\log(\lambda_i) \sim \alpha[\text{site}]$ [*the model*]

272 latency is the average latency to approach a novel environment, λ_i is the rate (probability of approaching
 273 the novel environment in each second) per bird (and we take the log of it to make sure it is always positive;
 274 birds with a higher rate have a smaller latency), ϕ is the dispersion of the rates across birds, and α is the
 275 intercept for the rate per site.

276 Expected values for the latency to approach a novel environment range from 0-2700 sec, which encompasses
 277 the time period during which they are exposed to the novel environment (sessions last up to 45 min).
 278 However, we do not provide an upper limit for the model because those birds that do not approach within
 279 2700 sec would eventually have had to approach the novel environment to access their food (it is just that
 280 sessions did not run that long). After running simulations, we identify the following distribution and priors
 281 to be the most likely for our expected data:

282 $\phi \sim 1/(\text{Exponential}(1))$ [ϕ prior]

283 $\alpha \sim \text{Normal}(1350, 500)$ [α prior]

284 We use a gamma-Poisson distribution for latency because it constrains the values to be positive. For ϕ , we
 285 used an exponential distribution because it is standard for this parameter. We used a normal distribution
 286 for α because it is a sum with a large mean (see Figure 10.6 in McElreath (2016)). We estimate that the
 287 grackles might approach the novel environment at any time in the session, therefore we hold the α mean of
 288 1350 sec in mind as we conduct the modeling. We set the α standard deviation to 500 because this puts the
 289 range of seconds for the distribution in the possible range.

290 Persistence analysis

291 Model and simulation

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

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

301 $\text{logit}(p) \sim \alpha[\text{site}]$ [*model*]

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

307 $\alpha \sim \text{Normal}(0, 0.5)$ [α prior]

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

311 Repeatability of exploration and persistence

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

320 Post-study choices made since receiving in principle recommendation

321 In the preregistration, we said that for the exploration measure we would use the “Latency to approach
 322 within 20 cm of an object (novel or familiar, that does not contain food) in a familiar environment (that
 323 contains maintenance diet away from the object) - OR - closest approach distance to the object (choose the
 324 variable with the most data for the analysis).” We had data for both exploration measures and we used the
 325 latency measure because this was the variable that our preregistered analysis was designed for.

326 In the peer review history of the preregistration, we said that we would use whichever exploration test was
 327 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
 328 and there was little variation in whether, or for how long, individuals went into the novel environment (i.e.,
 329 most individuals did not go in the novel environment). However, the Tempe grackles responded differently
 330 to the novel environment and novel object, therefore they did not perceive the stimuli as the same. From the
 331 Tempe grackle data, we found that responses were only repeatable for the novel environment test (McCune
 332 KB et al., 2019). Therefore, we conducted this assay (and not the novel object assay) with the Woodland
 333 grackles and compared the two populations on this one assay.

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

345 The exploration data for the repeatability calculation were heteroscedastic and overdispersed. Additionally,
 346 53% of the data were at the ceiling value (i.e., the bird did not approach the novel environment). Conse-
 347 quently, the model that best fit the data and was appropriate for the repeatability analysis was a binomial
 348 model, where the response was 0 (the grackle never approached the novel environment during exploration
 349 trials) or 1 (the grackle approached the novel environment).

350 **RESULTS**

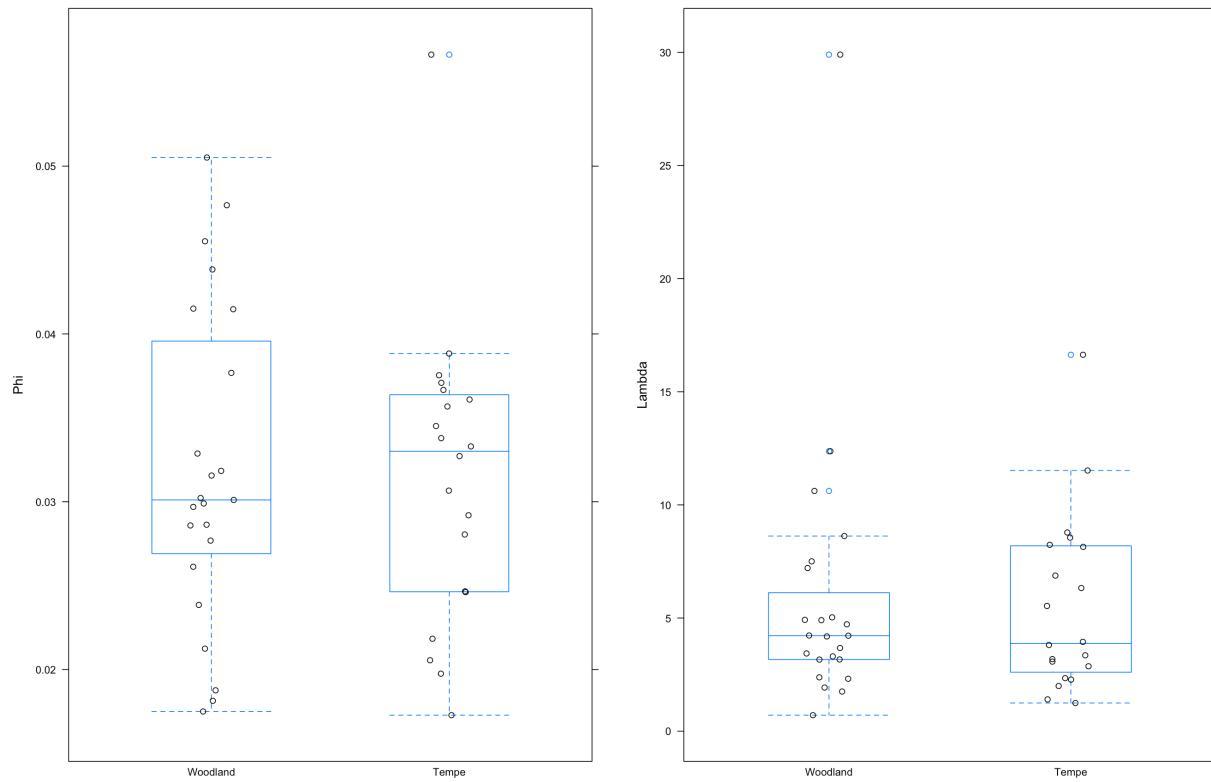
 351 **Flexibility**

352 There are no strong site differences for either component of reversal learning: ϕ or λ (Figure 3). However, ϕ
 353 differs by only 0.0005 (Woodland=0.0306, Tempe=0.0301) and λ by 0.26 (Woodland=4.78, Tempe=4.52),
 354 and the compatibility intervals for the estimated differences for both parameters cross zero (Table 2). With
 355 our sample size, we only have the power to reliably detect differences between the populations if they are
 356 larger than 0.01 for ϕ and 1.4 for λ (based on our power analysis in Supplementary Material 2, summarized in
 357 Table SM1). Accordingly, we cannot exclude that the two populations are different, however we can estimate
 358 the range for how small the difference can be. Based on the estimated 89% compatibility intervals for phi
 359 and lambda in Table 2, the two populations are unlikely to differ by more than 0.01 for ϕ and 3 for λ (Table
 360 2). Plotting the values (Figure 3) suggests no differences in the variances because similar minimum and
 361 maximum values are observed in both populations.

362 **Table 2.** Contrasts (indicated by “diff”) between populations for the flexibility measure of reversal learning:
 363 phi and lambda.

	Mean	Standard deviation	Lower 89 percentile compatibility interval (5.5%)	Upper 89 percentile compatibility interval (94.5%)
Woodland Phi	0.03	0.01	0.02	0.05
Woodland Lambda	5.84	5.96	1.79	12.00
Tempe Phi	0.03	0.01	0.02	0.04
Tempe Lambda	5.51	3.93	1.43	11.40
diff_Phi	0.00	0.01	-0.01	0.01
diff_Lambda	0.26	0.68	-0.73	1.40

365



366

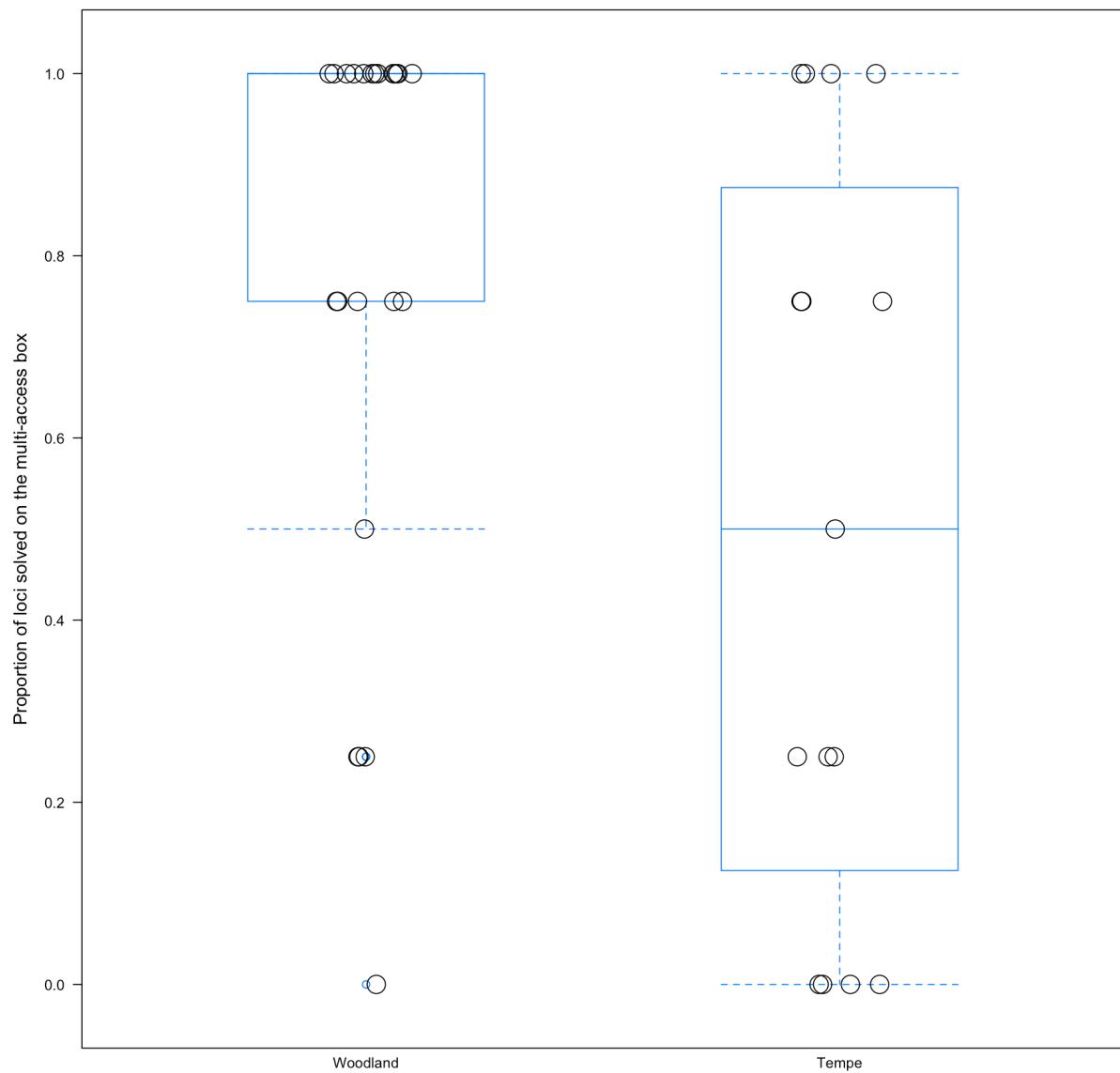
367 **Figure 3.** Measures of flexibility from the reversal learning experiment: ϕ and λ per individual in each
 368 population. The boxplots show the minimum, maximum, lower and upper quartiles, and median values.
 369 The blue circles are outliers associated with the boxplots. The black circles are the raw data from each
 370 individual.

371 **Innovation**

372 Individuals in the more recent population, Woodland, California, are more innovative than individuals in the
 373 older population in Tempe, Arizona (Figure 4). Woodland grackles solve a higher proportion of loci on the
 374 multiaccess box as indicated by the contrast that showed that the compatibility interval did not cross zero
 375 (diff_12 in Table 3). Plotting the values (Figure 4) suggests no clear differences in the variances between
 376 the two populations because some individuals in both populations solved zero and some solved all four loci.

377 **Table 3.** Contrasts between populations for the innovation measure: the proportion of loci solved on the
 378 multi-access box.

	Mean	Standard deviation	Lower 89 percentile compatibility interval (5.5%)	Upper 89 percentile compatibility interval (94.5%)
379 Woodland	0.76	0.04	0.69	0.83
Tempe	0.50	0.06	0.41	0.60
380 diff_12	0.26	0.07	0.14	0.37



381

382 **Figure 4.** Number of loci solved on the multiaccess box in the innovativeness test per individual at each
 383 site ($n=21$ birds in Woodland, $n=15$ birds in Tempe). The boxplots show the minimum, maximum, lower
 384 and upper quartiles, and median values. The blue circles are outliers associated with the boxplots. The
 385 black circles are the raw data from each individual.

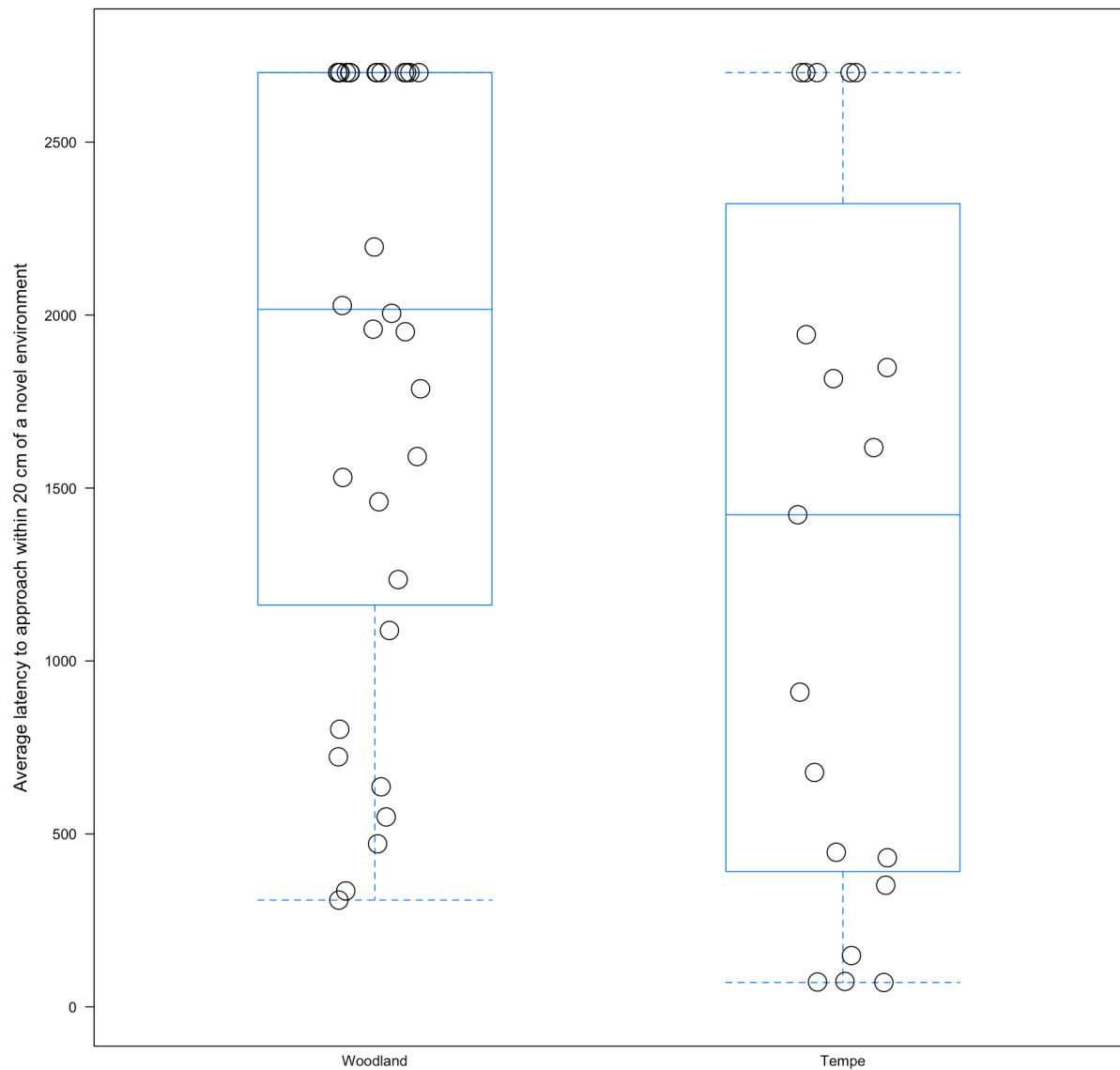
386 **Exploration**

387 Individuals in the older population, Tempe, Arizona, are more exploratory than individuals in the more
 388 recent population in Woodland, California (Figure 5). Tempe grackles are faster (have lower latencies) to
 389 approach a novel environment as indicated by the contrast that shows that the compatibility interval does
 390 not cross zero (diff_12 in Table 4). Plotting the values (Figure 5) suggest no clear differences in the variances
 391 between the two populations because there is a similar spread of latencies.

392 **Table 4.** Contrasts (indicated by “diff”) between populations for the exploration measure: latency to
 393 approach within 20 cm of a novel environment. Note that “phi” in this table refers to a term in the gamma

394 poisson model and not to what we refer to as the phi parameter in reversal learning.

	Mean	Standard deviation	Lower 89 percentile compatibility interval (5.5%)	Upper 89 percentile compatibility interval (94.5%)
395 Woodland	1697.40	229.76	1368.91	2058.43
Tempe	1137.56	181.84	875.64	1448.64
phi	1.59	0.29	1.15	2.09
396 diff_12	559.84	285.99	103.84	1017.56



397

398 **Figure 5.** Average latency to approach within 20 cm of a novel environment in the exploration assay per
 399 individual at each site (n=32 Woodland, n=19 Tempe). Note that if an individual does not approach within
 400 20 cm of the novel environment at Time 1 or 2, they are given a ceiling value of 2701, which is one second
 401 longer than the session length. The boxplots show the minimum, maximum, lower and upper quartiles, and

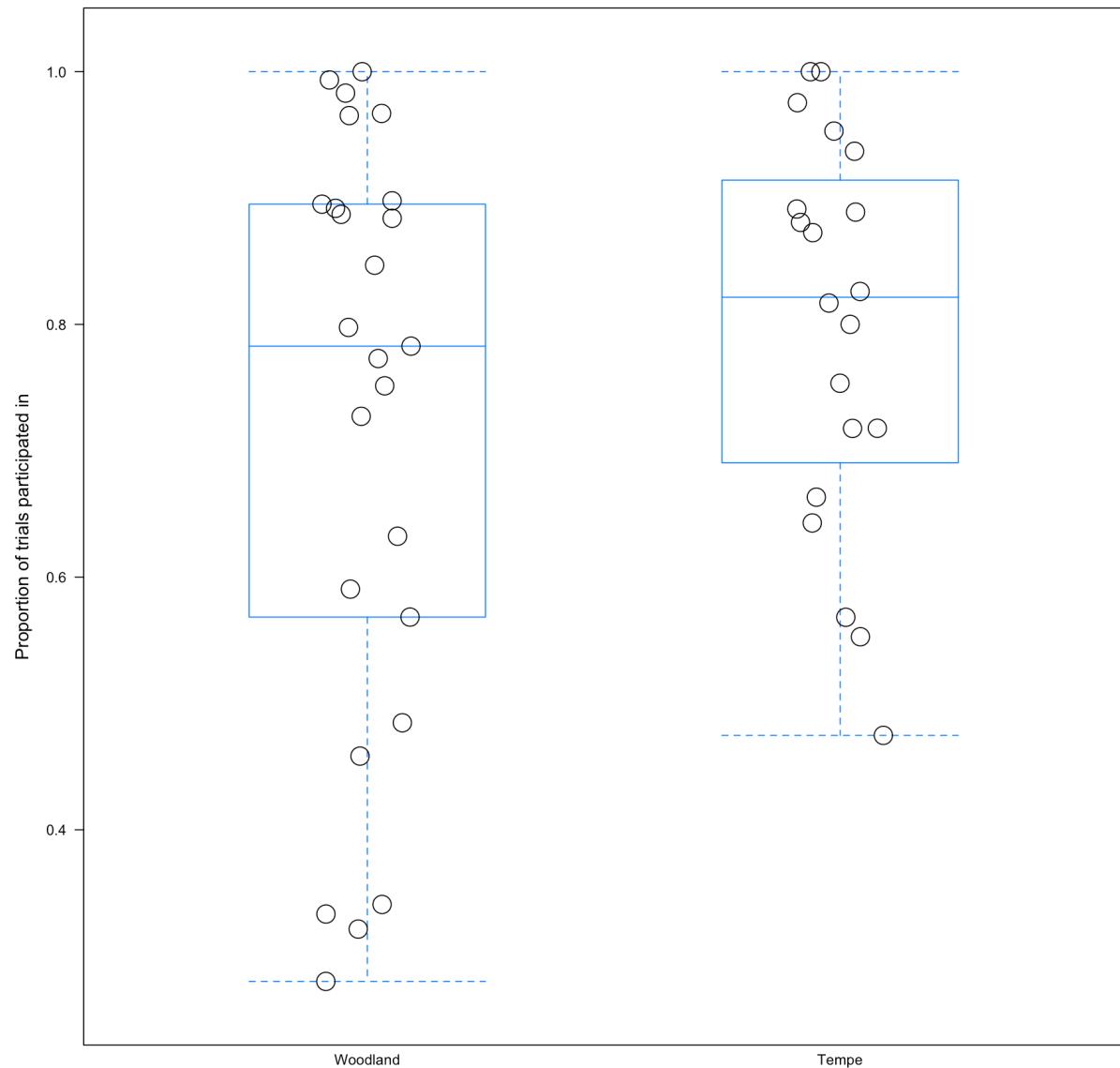
402 median values. The black circles are the raw data from each individual.

403 **Persistence**

404 There are no strong site differences for persistence quantified as the proportion of trials participated in across
 405 the reversal and multiaccess box experiments (Figure 6). We would need a difference of more than 0.08 in
 406 the proportion of trials participated in to detect a difference between the sites (based on our power analysis
 407 in Supplementary Material 2, summarized in Table SM1). However, the proportion differs by only 0.08
 408 (Woodland=0.72, Tempe=0.80), and the site differences are unlikely to be larger than 0.08 (Table 5). Visual
 409 interpretation, through plotting the values (Figure 6), could suggest that the variance in persistence might
 410 be larger among the individuals in Woodland compared to Tempe because some of the Woodland individuals
 411 show lower persistence values than those in the Tempe individuals. We conducted an UNREGISTERED
 412 ANALYSIS which finds no support that the variances differ between the two populations (Levene's test for
 413 homogeneity of variance: df=1, F value=1.9, p=0.17).

414 **Table 5.** Contrasts (indicated by “diff”) between populations for the persistence measure: proportion of
 415 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%)
416 Woodland	0.78	0.01	0.77	0.80
Tempe	0.79	0.01	0.78	0.80
417 diff_12	0.00	0.01	-0.02	0.01



418

419 **Figure 6.** The proportion of trials participated in across the reversal and multiaccess box experiments is
 420 the measure of persistence per individual at each site ($n=25$ Woodland, $n=20$ Tempe). The boxplots show
 421 the minimum, maximum, lower and upper quartiles, and median values. The black circles are the raw data
 422 from each individual.

423 **Repeatability of exploration and persistence**

424 Exploration of the novel environment is repeatable in the Woodland population (current study: likelihood
 425 ratio test: $R=0.70$, $p=0.001$, confidence interval=0.2-1.0). Our previous analysis found that novel envi-
 426 ronment exploration was repeatable in the Tempe (McCune KB et al., 2019: $R=0.72$, $p<0.001$, confi-
 427 dence interval=0.42-0.88) grackles. Persistence is repeatable across both populations (likelihood ratio test: $R=0.24$,
 428 $p=0.03$, confidence interval=0.03-0.46).

429 **DISCUSSION**

430 We conduct behavioral experiments with great-tailed grackles from two populations: an older population in
431 the middle of the expansion front (Tempe, Arizona), and a more recent population on the northern edge of
432 their expansion in Woodland, California. Our measures of flexibility (using serial reversals in the Tempe popu-
433 lation, K. McCune et al., 2022), exploration (Tempe, McCune KB et al., 2019, Woodland, reported here),
434 and persistence (both populations reported here) are repeatable and show large inter-individual variation,
435 which validates that these are stable traits that can be meaningfully compared. We find that individuals
436 in the edge population are more innovative and less exploratory than the population in the middle of the
437 expansion front, and that there are no population differences in behavioral flexibility or persistence. This
438 supports the hypothesis that changes in particular behavioral traits are potentially important for facilitating
439 a species' rapid geographic range expansion.

440 We find no support for the hypothesis that flexibility plays an important role in rapid geographic range
441 expansions (Chow et al., 2016; Griffin & Guez, 2014; e.g., Lefebvre et al., 1997; Sol et al., 2002, 2005, 2007;
442 Sol & Lefebvre, 2000; Wright et al., 2010). The finding that flexibility is not higher among individuals at the
443 edge of the expansion range indicates that flexibility is not a latent trait that is called upon when individuals
444 move into new areas.

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

474 The ability of great-tailed grackles to move into new habitats might be a species specific ability that has
475 been ongoing for many years, which could be linked to the high levels of flexibility in this species being
476 relatively fixed (Wright et al., 2010). great-tailed grackles are flexible on the reversal learning task and are
477 perhaps at their upper limit uniformly across their range. With an average reversal learning speed of 74
478 trials (using the data in the current article), great-tailed grackles are as flexible as great (*Parus major*) and
479 blue (*Cyanistes caeruleus*) tits [average 59 trials; Morand-Ferron et al. (2022)] and three species of Darwin's
480 finches (average 89 trials); and more flexible than Pinyon jays (average 155 trials), Clark's nutcrackers

481 (average 143 trials), California scrub jays (average 191 trials), pigeons (average 168 trials) Lissek et al.
482 (2002), and mice (average approximately 150 trials, Laughlin et al., 2011). Perhaps great-tailed grackles
483 maintain a high level of flexibility across their range in response to daily changes in their local environment
484 (e.g., the changing schedules of cafes with outdoor seating areas and garbage pick up times), rather than
485 specifically in response to larger changes that might occur less frequently (e.g., traveling farther to exploit
486 new foraging opportunities or moving to a new area).

487 Another alternative is that we measured the edge population too long after their initial establishment,
488 during which time they potentially exhibited more flexibility for their initial adaptation phase to the new
489 area (Wright et al., 2010). If the sampled individuals had already been living at this location for long enough
490 (or for their whole lives) to have learned what they need to about this particular environment (e.g., there
491 may no longer be evidence of increased flexibility/innovativeness/exploration/persistence), there may be no
492 reason to maintain population diversity in these traits to continue to learn about this environment. In this
493 case, because differences in innovativeness are found, this trait could have different timing in the process
494 of establishing in a new location (i.e., be required for longer). Great-tailed grackles occur more irregularly
495 in areas further north of our edge site, and flexibility might be higher in more northern individuals from
496 areas where stable populations are not yet established. However, evidence from experimental evolution
497 suggests that, even after 30 generations there is no change in exploration of a novel environment or other
498 behaviors (aggression, social grooming, courtship, and orientation) when comparing domestic guinea pigs
499 with 30 generations of wild-caught captive guinea pigs (Künzl et al., 2003), whereas artificial selection can
500 induce changes in spatial ability in as little as two generations (Kotrschal et al., 2013). This means it is
501 likely that we would have detected population differences if such differences were linked with adapting to a
502 new environment.

503 Differences in innovativeness and exploration are associated with the great-tailed grackle's rapid geographic
504 range expansion. An increase in innovation in newly established populations can facilitate innovating new
505 foraging techniques and exploiting new food sources, while a decrease in exploration can reduce their risk of
506 encountering danger in a new area. The relatively little evidence from invasive species that are also expanding
507 their geographic ranges shows similar results. Common mynas (*Acridotheres tristis*) on the invasion front
508 are more innovative than those from populations away from the front and in their native range (Cohen et al.,
509 2020), and spiders from edge populations are less exploratory than those from core populations (Chuang &
510 Riechert, 2021). While great-tailed grackles are not considered an invasive species because they introduced
511 themselves rather than being introduced by humans, comparing them with invasive species is useful because
512 the dynamics after the introduction stage should be similar (i.e., establishing in a new area and spreading
513 out from there) (Chapple et al., 2012). Note that wild great-tailed grackles were caught from north of Rio
514 de la Antigua, Mexico by the Aztec emperor, Auitzotl (1486-1502), and introduced approximately 370 km
515 inland to the Valley of Mexico (Tenochtitlan & Tlatelolco) where they reproduced and spread (P. Haemig,
516 2014; P. D. Haemig, 2011, 2012). By 1577, they spread at least 100 km including back to their native
517 range (P. D. Haemig, 2011). This indicates that great-tailed grackles had already spread this far north by
518 themselves before the introduction at a parallel latitude, and that they continued their spread without the
519 help of human-facilitated introductions.

520 Flexibility is causally related with innovativeness in great-tailed grackles (Logan et al., 2023, measured on
521 the Tempe individuals included in the current study). We manipulated flexibility in the Tempe grackles
522 by giving a manipulated group serial reversals until they passed quickly. The manipulated grackles were
523 then given an innovation test (the multiaccess box) and found to be more innovative (solved more loci)
524 compared to control grackles who only experienced one reversal. Flexibility, the ability to recognize that
525 something about the environment has changed and decide to consider other options for deploying behavior
526 (Mikhalevich et al., 2017), is distinct from innovation, which is the specific stringing together of particular
527 behaviors in response to the decision to change behavior in some way (Griffin & Guez, 2014). That they are
528 causally related does not mean that they must always be associated to the same degree because there can
529 be other variables that additionally influence one or both traits differentially across time and space (e.g.,
530 environmental unpredictability, features of the items they forage on that differ and require different access
531 methods). We are currently investigating how flexibility and exploration, and flexibility and a different
532 measure of persistence (number of functional and/or non-functional touches to test apparatuses) are related
533 in the Tempe grackles (McCune KB et al., 2019). Additionally, we are determining to what extent the aviary

534 measure of exploration is a proxy for how the Tempe individuals use space in the wild after their release
535 (McCune KB et al., 2020).

536 In conclusion, rather than flexibility being associated with a rapid geographic range expansion, as is widely
537 hypothesized, we find that higher innovation and lower exploration levels are the key behavioral traits
538 associated with the great-tailed grackle's edge population in comparison with an older population closer
539 to the original range. The term "behavioral flexibility" is defined and measured in a variety of ways in
540 the literature (or it is not defined at all). For example, the detour task (individuals must walk around a
541 transparent barrier to access a food reward) is sometimes considered a test of flexibility, sometimes a test
542 of self control, and sometimes a test of both. However, theoretically and empirically it measures a trait
543 that is not, and is not related to, flexibility or self control, but rather a different trait: motor inhibition
544 (Logan et al., 2021). We argue that calling many types of traits "flexibility" without proper (or sometimes
545 any) theoretical justification and without validating methods is detrimental because it confounds our ability
546 to answer questions about the broader significance of flexibility and how it is genuinely involved in large
547 scale changes (Logan et al., 2017; Mikhalevich et al., 2017). Our research program shows the value of
548 clearly defining terms for behavioral traits, validating the methods intended to measure those traits, and
549 understanding how certain traits relate to each other (causally if possible) before attempting to answer
550 broader cross population questions.

551 ETHICS

552 This research was carried out in accordance with permits from the:

- 553 1) US Fish and Wildlife Service (scientific collecting permit number MB76700A-0,1,2)
- 554 2) US Geological Survey Bird Banding Laboratory (federal bird banding permit number 23872)
- 555 3) Arizona Game and Fish Department (scientific collecting license number SP594338 [2017], SP606267
556 [2018], SP639866 [2019], and SP402153 [2020])
- 557 4) Institutional Animal Care and Use Committee at Arizona State University (protocol number 17-1594R)
- 558 5) California Department of Fish and Wildlife (scientific collecting permit [specific use] number
559 S-192100001-19210-001)
- 560 6) RegionalSan (access permit number AP 2021-01)

561 AUTHOR CONTRIBUTIONS

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

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

565 **LeGrande-Rolls:** Data collection, revising/editing.

566 **Marfori:** Data collection, revising/editing.

567 **Hubbard:** Data collection, revising/editing.

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

569 FUNDING

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

572 **CONFLICT OF INTEREST DISCLOSURE**

573 We, the authors, declare that we have no financial conflicts of interest with the content of this article. CJ
574 Logan and D Lukas are Recommenders at PCI Ecology, and CJ Logan used to be on the Managing Board
575 at PCI Ecology (2018-2022).

576 **ACKNOWLEDGEMENTS**

577 We thank Kristine Johnson for technical advice on great-tailed grackles; Julia Cissewski and Sophie Kaube
578 for tirelessly solving problems involving financial transactions and contracts; Richard McElreath for project
579 support; Aaron Blackwell and Ken Kosik for being the UCSB sponsors of the Cooperation Agreement with
580 the Max Planck Institute for Evolutionary Anthropology; Alexis Breen and Vincent Kiepsch for video coding;
581 Ann Brice, Woodland-Davis Clean Water Agency, Regional San, and Conaway Ranch for hosting the research
582 on their land; and Rhonda Oates and the vet team at UC Davis for veterinary consultations.

583 **SUPPLEMENTARY MATERIAL 1: Sample size rationale**

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

587 **Table SM1.** A summary of the measure of interest in each experiment, the distribution used for the analysis,
 588 the minimum detectable difference between site means, and the minimum sample size that goes with the
 589 minimum detectable difference.

Experiment	Measurement	Distribution	Minimum difference between site means	Minimum sample size
Reversal	Phi (learning rate)	Gamma	Differences of 0.01 are likely to be detected (based on models with 20 individuals per site, however this is likely to hold for the the minimum sample size as well) (Figures 5 and 6)	15
Reversal	Lambda (random choice rate)	Gamma	Differences of 3 are likely to be detected (based on models with 20 individuals per site, however this is likely to hold for the the minimum sample size as well) (Figures 5 and 6)	15
590				
Multiaccess box	Number of loci solved	Binomial	Differences of 1.2 loci are likely to be detected (Table 3)	15
Exploration	Latency to approach novel object	Gamma-Poisson	Differences of at least 450 seconds are likely to be detected (Table 3)	15
Persistence	Percent of trials participated in	Normal	Difference of at least 0.08 in the proportion of trials participated in (Table 3)	18

591 **SUPPLEMENTARY MATERIAL 2: Simulations for power analyses**

592 **Hypothesis-specific mathematical model**

593 Following procedures in McElreath (2016), we construct a **hypothesis-appropriate mathematical model**
 594 for each of the response variables that examines differences in the response variable between sites (each site
 595 represents a grackle population). These models take the form of:

596 $y \sim \alpha[\text{site}]$

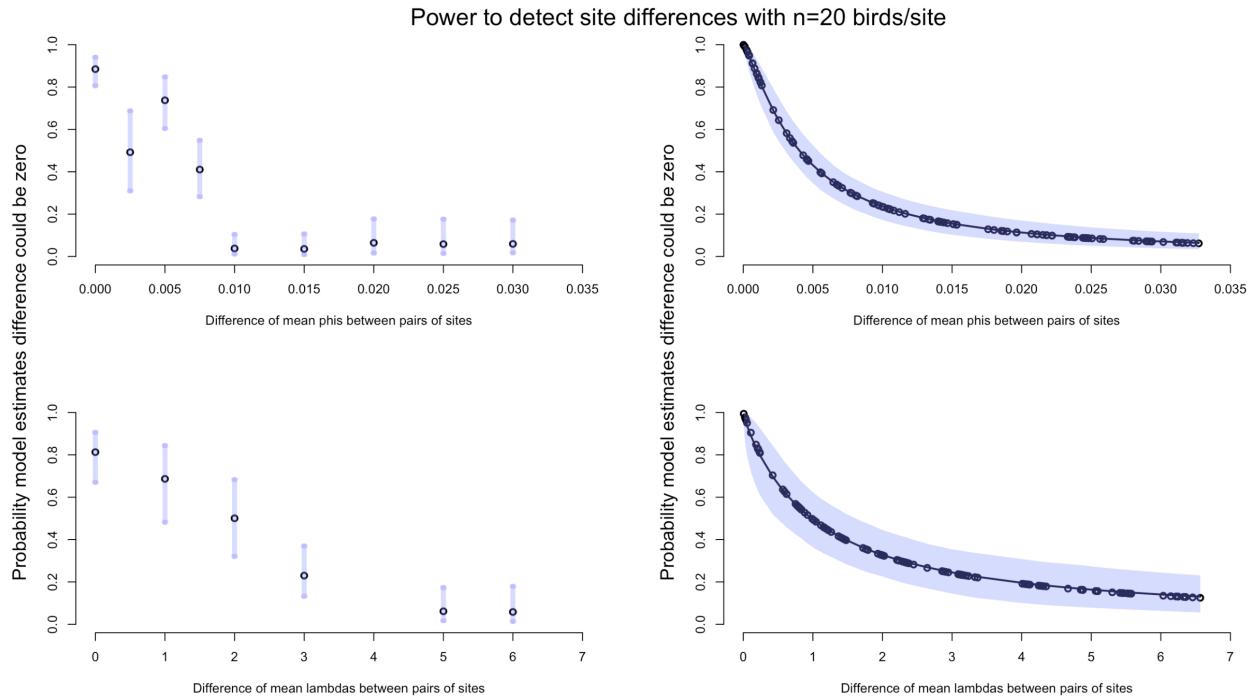
597 y is the response variable (flexibility, innovation, exploration, or persistence). There is one intercept, α , per
 598 site and we estimate the site's average and standard deviation of the response variable.

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

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

617 **Flexibility analysis**

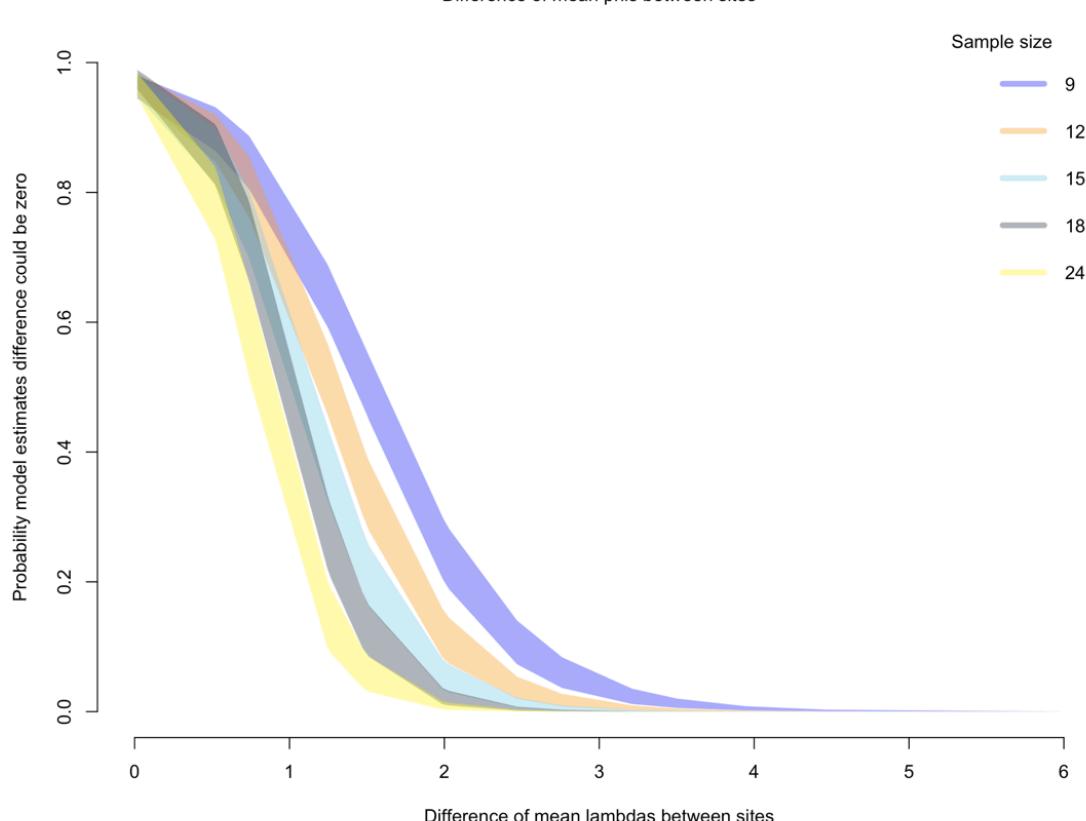
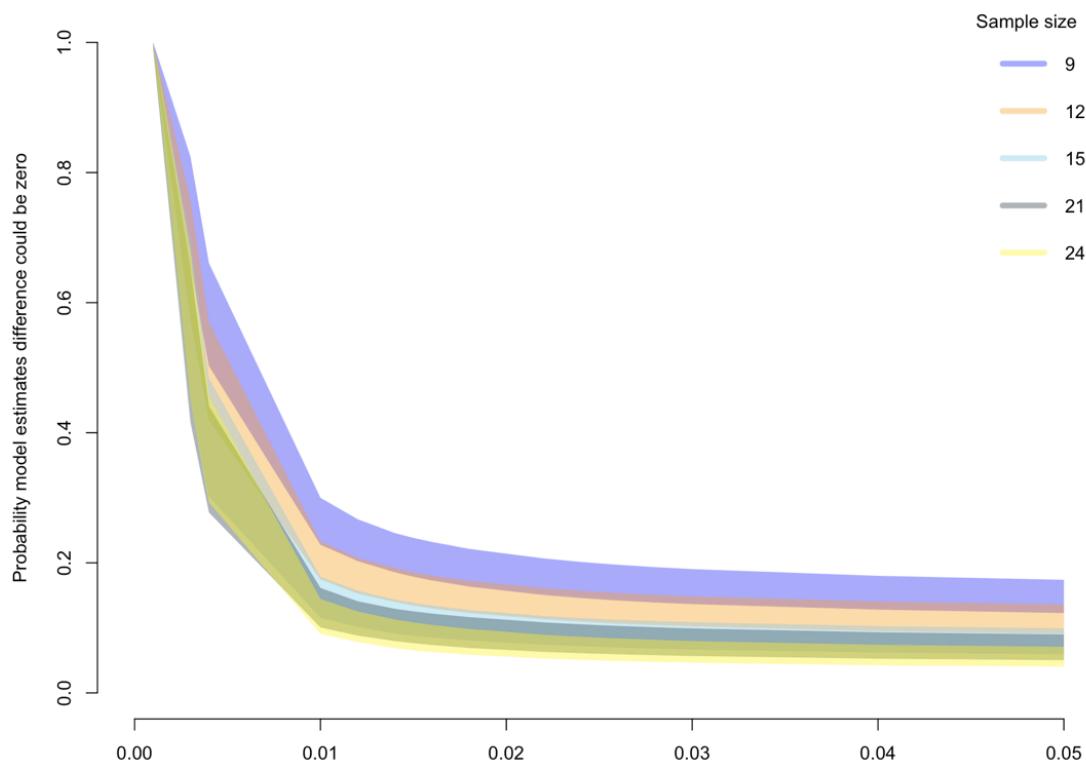
618 **Power analyses:** We also use the simulations to estimate our ability to detect differences in ϕ and λ between
 619 sites based on extracting samples from the posterior distribution. We run two different sets of simulations:
 620 we first sample between 9 and 24 birds from populations with pre-specified ϕ and λ means to determine
 621 the minimum sample size required to detect whether two populations are different. This set of simulations
 622 shows how different site sample sizes change detection levels: once a sample size of 15 is reached, there are
 623 only minimal differences in detection abilities compared to larger sample sizes (Figure SM2.1). The second
 624 set of simulations recreates choices for 20 birds per population across initial learning and reversal trials from
 625 which we estimate their ϕ and λ . We simulate 20 birds per population because this number is above the
 626 threshold we detected in the first set of simulations and it appears a feasible sample size. We expect that the
 627 noise in the probabilistic choices of individuals might reduce the differences that can be detected compared
 628 to the first simulation where ϕ and λ are assumed to be exactly known for each individual. This second
 629 set of simulations shows that we have a very high chance of detecting that two sites are different from each
 630 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
 631 on data from 20 simulated individuals per site (Figure SM2.2). It appears that there is more variability in
 632 the λ estimates for each bird based on their choices, meaning that with the learning model, which estimates
 633 λ from the choices, the differences between sites have to be larger (than if we were able to infer lambda
 634 directly) to be reliably detected. The power curves in Figure SM2.1 are more reliable than those in Figure
 635 2.2.



636

637 **Figure SM2.1.** How small of a site difference in phi and lambda can we detect? The probability that
 638 the model estimates that the difference shown on the x axis is zero, meaning that the model assumes
 639 that it is possible that these two estimates come from a population with the same phi or lambda. Each
 640 point is the mean phi or mean lambda from one site minus the mean phi or mean lambda from another
 641 site (calculated from 20 individuals per site) for all pairwise comparisons for all 32 sites (for a total of
 642 496 pairwise comparisons). Left panels: error bars=89% compatibility intervals. Right panels: shaded
 643 areas=97% prediction intervals.

Power to detect site differences



645 **Figure SM2.2.** How do detection differences vary according to sample size differences? The probability
 646 that the model estimates that the difference shown on the x axis is zero, meaning that the model assumes
 647 that it is possible that these two estimates come from a population with the same phi or lambda. The x-axis
 648 is the mean phi or mean lambda from one site minus the mean phi or mean lambda from another site for
 649 all pairwise comparisons for all 14 sites (for a total of 91 pairwise comparisons). Each shaded region is the
 650 97% prediction interval for that particular sample size.

651 **Innovation analysis**

652 After building the model (see Methods), we then run the **mathematical model** and perform pairwise
 653 contrasts and determine that we are able to detect differences between sites with a sample size of 15 at each
 654 site if the average number of loci solved differs by 1.2 loci or more and the standard deviation is generally
 655 a maximum of 0.9 at each site (Table SM2). For a sample size of 20 at each site, we are able to detect site
 656 differences if the average number of loci solved differs by 0.7 of a locus or more and the standard deviation is
 657 generally a maximum of 1 at each site (Table SM2). Note: the Arizona sample size is 11 for the multiaccess
 658 log and 17 on a similar multiaccess box.

659 **Table SM2.** Simulation outputs from **varying sample size (n), and α means and standard deviations**. We calculate pairwise contrasts between the estimated means from the posterior distribution: if for
 660 a large sample the difference is both positive and negative and crosses zero (yes), then we are not able to
 661 detect differences between the two sites. If the differences between the means are all on one side of zero
 662 for 89% of the posterior samples (no), then we are able to detect differences between the two sites. We
 663 chose the 89% interval based on (McElreath, 2016). Note that for latency, there is no mu_sd, but rather
 664 one phi that is the same for all sites. The numbers 1-3 in the column titles refer to sites 1-3 as do S1-3
 665 (the simulations were run on a total of three sites because we originally planned to collect data at two
 666 to three sites), mu=average, sd=standard deviation. Loci solved is the innovativeness measure, latency
 667 is the exploration measure, and trials participated in is the persistence measure. Note that the full table
 668 can be viewed at https://raw.githubusercontent.com/corinalogan/grackles/master/Files/Preregistrations/gxpopbehaviorhabitat_table_simoutputs.csv
 670

672

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

677 Exploration analysis

678 After building the model (see Methods), we then run the **mathematical model** and perform pairwise
679 contrasts and determine that we are able to detect differences between sites with a sample size of 15 or 20 at
680 each site if the average latency to approach the novel environment differs by at least 450 sec between sites
681 (Table SM2). We keep the shape of the curve (which can be thought of as similar to a standard deviation or
682 the variance) the same across sites because we do not think this assumption will change across populations
683 (i.e., there will be lots of variation at each site with some individuals approaching almost immediately, others
684 in the middle of the session, and others near the end).

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

689 Persistence analysis

690 After building the model (see Methods), we then run the **mathematical model** and perform pairwise
691 contrasts and determine that we are able to detect differences between sites with a sample size of 15 or 20
692 per site if the average proportion of trials participated in differs by at least 0.08 and the standard deviation
693 is generally a maximum of 0.25 at each site (Table 2).

694 Because the mean and the variance are linked in the binomial distribution, and because the variance sim-
695 ulations in the flexibility analysis show that we are not able to robustly detect differences in variance between
696 sites, we plot the variance in the proportion of trials participated in between sites to determine whether the
697 edge population has a wider or narrower spread than the other two populations.

698 **SUPPLEMENTARY MATERIAL 3: Interobserver reliability of dependent vari-
699 ables**

700 To determine whether experimenters coded the dependent variables in a repeatable way, hypothesis-blind
701 video coders were first trained in video coding the dependent variables (reversal learning and multiaccess
702 log: whether the bird made the correct choice or not; exploration: latency to approach), requiring a Cohen's
703 unweighted kappa (reversal and multiaccess categorical variables) or an intra-class correlation coefficient
704 (ICC; exploration continuous variable) of 0.90 or above to pass training. This threshold indicated that
705 the two coders (the experimenter and the video coder) agreed with each other to a high degree (kappa:
706 Landis & Koch (1977); ICC: Hutcheon et al. (2010)). After passing training, the video coders coded 20%
707 of the videos for each experiment (except for exploration for which 15% of the videos were coded due to
708 an unexpectedly high sample size for this assay). The kappa and ICC were calculated to determine how
709 objective and repeatable scoring was for each variable, while noting that the experimenter has the advantage
710 over the video coder because watching the videos is not as clear as watching the bird participate in the
711 trial from the aisle of the aviaries. The unweighted kappa was used when analyzing a categorical variable
712 where the distances between the numbers are meaningless (0=incorrect choice, 1=correct choice, -1=did not
713 participate), and the ICC was used for continuous variables where distances are meaningful (e.g., if coders
714 disagree by a difference of 2 s rather than 5 s, this is important to account for).

715 **Interobserver reliability training**

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

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

- 722 • Reversal learning: correct choice unweighted Cohen's Kappa=1.00 (confidence boundaries=1.00-1.00,
723 n=21 data points)
- 724 • Multiaccess box: correct choice unweighted Cohen's Kappa=1.00 (confidence boundaries=1.00-1.00,
725 n=29 data points)
- 726 • Multiaccess box: correct choice unweighted Cohen's Kappa=1.00 (confidence boundaries=1.00-1.00,
727 n=29 data points)

728 **Interobserver reliability**

729 Interobserver reliability scores (minimum 15% of the videos) were as follows:

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

- 731 • Reversal learning (5/19 birds): correct choice unweighted Cohen's Kappa=1.00 (confidence
732 boundaries=0.99-1.00, n=707 data points)
- 733 • Multiaccess box (5/23 birds): correct choice unweighted Cohen's Kappa=0.92 (confidence
734 boundaries=0.81-1.00, n=63 data points)
- 735 • Multiaccess box (5/23 birds): locus solved unweighted Cohen's Kappa=1.00 (confidence boundaries=1.00-
736 1.00, n=48 data points)

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



- 738 • Exploration (5/34 birds): latency to land on the ground unweighted Cohen's Kappa=0.998 (confidence
739 boundaries=0.997-0.999, n=32 data points)

740 **SUPPLEMENTARY MATERIAL 4: Boat-tailed grackle model outputs**

741 **Table SM4.** Results for the comparison between the boat-tailed grackle (BTGR) population in Lake Placid
 742 and Venus, Florida and the great-tailed grackle populations in Tempe, Arizona and Woodland, California.
 743 Contrasts (indicated by “diff”) between populations show whether there was a difference (compatibility
 744 interval does not cross zero) or not (compatibility interval crosses zero) for that pair of populations. Popu-
 745 lations are labeled as follows: 1=boat-tailed grackles (BTGR), 2=Woodland great-tailed grackles, 3=Tempe
 746 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	0.03	0.01	0.02	0.04
BTGR lambda	4.51	1.34	3.11	5.93
diff_12 phi	0.00	0.01	-0.01	0.01
diff_12 lambda	0.23	0.97	-1.06	1.97
diff_13 phi	0.00	0.01	-0.01	0.01
diff_13 lambda	0.43	1.01	-0.79	2.32
	NA	NA	NA	NA
INNOVATIVENESS	NA	NA	NA	NA
BTGR	0.36	0.11	0.19	0.53
Woodland	0.76	0.04	0.69	0.83
Tempe	0.50	0.06	0.40	0.60
diff_12	-0.41	0.12	-0.59	-0.22
diff_13	-0.14	0.13	-0.34	0.07
	NA	NA	NA	NA
PERSISTENCE	NA	NA	NA	NA
BTGR	0.69	0.02	0.66	0.72
Woodland	0.78	0.01	0.77	0.79
Tempe	0.79	0.01	0.78	0.80
diff_12	-0.10	0.02	-0.13	-0.06
diff_13	-0.10	0.02	-0.13	-0.06

748

749 REFERENCES

- 750 Auersperg, AMI., Bayern, A. M. P. von, Gajdon, G. K., Huber, L., & Kacelnik, A. (2011). Flexibility in
 751 problem solving and tool use of kea and New Caledonian crows in a multi access box paradigm. *PLOS
 752 ONE*, 6(6), e20231. <https://doi.org/10.1371/journal.pone.0020231>
- 753 Auersperg, A., Szabo, B., Von Bayern, A. M., & Kacelnik, A. (2012). Spontaneous innovation in tool
 754 manufacture and use in a goffin's cockatoo. *Current Biology*, 22(21), R903–R904.
- 755 Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4.
 756 *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- 757 Bates, D., Maechler, M., & Bolker, B. (2012). *lme4: Linear mixed-effects models using S4 classes* (2011). *R
 758 package version 0.999375-42*.
- 759 Bird, C. D., & Emery, N. J. (2009). Insightful problem solving and creative tool modification by captive
 760 nontool-using rooks. *Proceedings of the National Academy of Sciences*, 106(25), 10370–10375.
- 761 BirdLife International. (2018). Quiscalus mexicanus. *The IUCN Red List of Threatened Species 2018*,
 762 e.T22724308A132174807. <http://dx.doi.org/10.2305/IUCN.UK.2018-2.RLTS.T22724308A132174807.en>
- 763 Blaisdell, A., Seitz, B., Rowney, C., Folsom, M., MacPherson, M., Deffner, D., & Logan, C. J. (2021). *Do
 764 the more flexible individuals rely more on causal cognition? Observation versus intervention in causal
 765 inference in great-tailed grackles (version 5 of this preprint has been peer reviewed and recommended by
 766 peer community in ecology [https://doi.org/10.24072/pci.ecology.100076])*. <https://doi.org/10.31234/osf.io/z4p6s>
- 767 Bolker, B. M. (2008). *Ecological models and data in r*. Princeton University Press.
- 768 Bond, A. B., Kamil, A. C., & Balda, R. P. (2007). Serial reversal learning and the evolution of behavioral
 769 flexibility in three species of north american corvids (*gymnorhinus cyanocephalus*, *nucifraga columbiana*,
 770 *aphelocoma californica*). *Journal of Comparative Psychology*, 121(4), 372.
- 771 Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo,
 772 J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical
 773 Software*, 76(1).
- 774 Chapple, D. G., Simmonds, S. M., & Wong, B. B. (2012). Can behavioral and personality traits influence
 775 the success of unintentional species introductions? *Trends in Ecology & Evolution*, 27(1), 57–64. <https://doi.org/10.1016/j.tree.2011.09.010>
- 776 Chow, P. K. Y., Lea, S. E., & Leaver, L. A. (2016). How practice makes perfect: The role of persistence,
 777 flexibility and learning in problem-solving efficiency. *Animal Behaviour*, 112, 273–283. <https://doi.org/10.1016/j.anbehav.2015.11.014>
- 778 Christensen, A. F. (2000). The fifteenth-and twentieth-century colonization of the basin of mexico by the
 779 great-tailed grackle (*quiscalus mexicanus*). *Global Ecology and Biogeography*, 9(5), 415–420.
- 780 Chuang, A., & Riechert, S. E. (2021). Personality shifts are inconsistent across two rapid range expansions
 781 of a non-native spider. *Animal Behaviour*, 173, 67–79. <https://doi.org/10.1016/j.anbehav.2020.12.018>
- 782 Cohen, T. M., Kumar, R. S., Nair, M., Hauber, M. E., & Dor, R. (2020). Innovation and decreased
 783 neophobia drive invasion success in a widespread avian invader. *Animal Behaviour*, 163, 61–72. <https://doi.org/10.1016/j.anbehav.2020.02.012>
- 784 Collias, E. C., & Collias, N. E. (1964). The development of nest-building behavior in a weaverbird. *The
 785 Auk*, 81(1), 42–52.
- 786 Eddelbuettel, D., & François, R. (2011). Rcpp: Seamless R and C++ integration. *Journal of Statistical
 787 Software*, 40(8), 1–18. <https://doi.org/10.18637/jss.v040.i08>
- 788 Gabry, J., & Češnovar, R. (2021). *Cmdstanr: R interface to 'CmdStan'*.
- 789 Gamer, M., Lemon, J., Gamer, M. M., Robinson, A., & Kendall's, W. (2012). Package “irr.” *Various
 790 Coefficients of Interrater Reliability and Agreement*.
- 791 Griffin, A. S., & Guez, D. (2014). Innovation and problem solving: A review of common mechanisms.
 792 *Behavioural Processes*, 109, 121–134. <https://doi.org/10.1016/j.beproc.2014.08.027>
- 793 Hadfield, J. (2010). MCMCglmm: Markov chain monte carlo methods for generalised linear mixed models.
 794 *Tutorial for MCMCglmm Package in R*, 125.
- 795 Hadfield, J. (2014). *MCMCglmm course notes*. <http://cran.r-project.org/web/packages/MCMCglmm/vignettes/CourseNotes.pdf>
- 796 Haemig, P. (2014). Aztec introduction of the great-tailed grackle in ancient mesoamerica: Formal defense of

- 802 the sahaguntine historical account. *NeoBiota*, 22, 59–75. <https://doi.org/10.3897/neobiota.22.6791>
- 803 Haemig, P. D. (2011). Introduction of the great-tailed grackle by aztec emperor auitzotl: Four-stage analysis
 804 with new information. *Ardeola*, 58(2), 387–397. <https://doi.org/10.13157/arla.58.2.2011.387>
- 805 Haemig, P. D. (2012). Introduction of the great-tailed grackle (*Quiscalus mexicanus*) by aztec emperor
 806 auitzotl: Provenance of the historical account. *The Auk*, 129(1), 70–75. <https://doi.org/10.1525/auk.2011.11058>
- 807 Hartig, F. (2019). *DHARMA: Residual diagnostics for hierarchical (multi-level / mixed) regression models*.
 808 <http://florianhartig.github.io/DHARMA/>
- 809 Hutcheon, J. A., Chiolero, A., & Hanley, J. A. (2010). Random measurement error and regression dilution
 810 bias. *Bmj*, 340, c2289. <https://doi.org/10.1136/bmj.c2289>
- 811 Johnson, K., & Peer, B. D. (2001). *Great-tailed grackle: Quiscalus mexicanus*. Birds of North America,
 812 Incorporated.
- 813 Kotrschal, A., Rogell, B., Bundsen, A., Svensson, B., Zajitschek, S., Brännström, I., Immler, S., Maklakov,
 814 A. A., & Kolm, N. (2013). Artificial selection on relative brain size in the guppy reveals costs and benefits
 815 of evolving a larger brain. *Current Biology*, 23(2), 168–171.
- 816 Künzl, C., Kaiser, S., Meier, E., & Sachser, N. (2003). Is a wild mammal kept and reared in captivity still
 817 a wild animal? *Hormones and Behavior*, 43(1), 187–196.
- 818 Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biomet-
 819 rics*, 159–174.
- 820 Laughlin, R. E., Grant, T. L., Williams, R. W., & Jentsch, J. D. (2011). Genetic dissection of behavioral
 821 flexibility: Reversal learning in mice. *Biological Psychiatry*, 69(11), 1109–1116.
- 822 Laumer, I., Call, J., Bugnyar, T., & Auersperg, A. (2018). Spontaneous innovation of hook-bending and
 823 unbending in orangutans (*Pongo abelii*). *Scientific Reports*, 8(1), 1–13.
- 824 Lefebvre, L., Whittle, P., Lascaris, E., & Finkelstein, A. (1997). Feeding innovations and forebrain size in
 825 birds. *Animal Behaviour*, 53(3), 549–560. <https://doi.org/10.1006/anbe.1996.0330>
- 826 Lissek, S., Diekamp, B., & Güntürkün, O. (2002). Impaired learning of a color reversal task after NMDA
 827 receptor blockade in the pigeon (*Columba livia*) associative forebrain (neostriatum caudolaterale). *Be-
 828 havioral Neuroscience*, 116(4), 523.
- 829 Logan, C. J. (2016a). Behavioral flexibility and problem solving in an invasive bird. *PeerJ*, 4, e1975.
 830 <https://doi.org/10.7717/peerj.1975>
- 831 Logan, C. J. (2016b). Behavioral flexibility in an invasive bird is independent of other behaviors. *PeerJ*, 4,
 832 e2215.
- 833 Logan, C. J., Avin, S., Boogert, N., Buskell, A., Cross, F. R., Currie, A., Jelbert, S., Lukas, D., Mares, R.,
 834 Navarrete, A. F., et al. (2018). Beyond brain size: Uncovering the neural correlates of behavioral and
 835 cognitive specialization. *Comparative Cognition & Behavior Reviews*.
- 836 Logan, C. J., Avin, S., Boogert, N., Buskell, A., Cross, F. R., Currie, A., Jelbert, S., Lukas, D., Mares, R.,
 837 Navarrete, A. F., Shigeno, S., & Montgomery, S. H. (2017). Beyond Brain Size. *bioRxiv*, 145334.
 838 <https://doi.org/10.1101/145334>
- 839 Logan, C. J., MacPherson, M., Rowney, C., Bergeron, L., Seitz, B., Blaisdell, A., Folsom, M., Johnson-
 840 Ulrich, Z., & McCune, K. B. (2023). Is behavioral flexibility manipulable and, if so, does it improve
 841 flexibility and innovativeness in a new context? *EcoEvoRxiv*. <https://doi.org/10.32942/osf.io/5z8xs>
- 842 Logan, C. J., McCune, K., MacPherson, M., Johnson-Ulrich, Z., Rowney, C., Seitz, B., Blaisdell, A., Deffner,
 843 D., & Wascher, C. (2021). *Are the more flexible great-tailed grackles also better at behavioral inhibition?*
 844 <https://doi.org/10.31234/osf.io/vpc39>
- 845 Logan, C. J., Shaw, R., Lukas, D., & McCune, K. B. (2022). *How to succeed in human modified environments*.
 846 <http://corinalogan.com/ManyIndividuals/mi1.html>
- 847 Logan, CJ, McCune, KB, Chen, N, & Lukas, D. (2020). Implementing a rapid geographic range expansion
 848 - the role of behavior and habitat changes. *In Principle Acceptance by PCI Ecology of the Version on 6
 849 Oct 2020*. <http://corinalogan.com/Preregistrations/gxpopbehaviorhabitat.html>
- 850 Logan, CJ, McCune, KB, Marfori Z, Rolls C, & Lukas, D. (2023). Data: Implementing a rapid geographic
 851 range expansion - the role of behavior changes. *Knowledge Network for Biocomplexity, Data package*.
 852 <https://doi.org/10.5063/F1QZ28FH>
- 853 Logan, C., Lukas, D., Blaisdell, A., Johnson-Ulrich, Z., MacPherson, M., Seitz, B., Sevchik, A., & McCune,
 854 K. (2023). Data: Is behavioral flexibility manipulatable and, if so, does it improve flexibility and problem
 855

- 856 solving in a new context? *Knowledge Network for Biocomplexity, Data package.* [https://doi.org/10.5063/
 857 F1XP73CJ](https://doi.org/10.5063/F1XP73CJ)
- 858 Manrique, H. M., & Call, J. (2011). Spontaneous use of tools as straws in great apes. *Animal Cognition*,
 859 14(2), 213–226.
- 860 McCune, K. B. (2018). *Cognition gone wild: A test of the social intelligence hypothesis in wild birds* [PhD
 861 thesis].
- 862 McCune, K. B., Jablonski, P., Lee, S., & Ha, R. R. (2019). Captive jays exhibit reduced problem-solving
 863 performance compared to wild conspecifics. *Royal Society Open Science*, 6(1), 181311.
- 864 McCune, KB, Folsom, M, Ross, C, Bergeron, L, & Logan, C. (2020). Does great-tailed grackle space use
 865 behavior reflect individual differences in exploration? In *Principle Acceptance by PCI Ecology of the*
 866 *Version on 23 Sep 2020.* <http://corinalogan.com/Preregistrations/gspaceuse.html>
- 867 McCune, K., Blaisdell, A., Johnson-Ulrich, Z., Lukas, D., MacPherson, M., Seitz, B., Sevchik, A., & Logan,
 868 C. (2022). Repeatability of performance within and across contexts measuring behavioral flexibility.
 869 *EcoEvoRxiv.* <https://doi.org/10.32942/osf.io/kevqp>
- 870 McCune, KB, MacPherson, M, Rowney, C, Bergeron, L, Folsom, M, & Logan, C. (2019). Is behavioral flexi-
 871 bility linked with exploration, but not boldness, persistence, or motor diversity? In *Principle Acceptance*
 872 *by PCI Ecology of the Version on 27 Mar 2019.* http://corinalogan.com/Preregistrations/g_exploration.html
- 873 McElreath, R. (2016). *Statistical rethinking: A bayesian course with examples in r and stan.* CRC Press.
 874 <https://doi.org/10.1201/9781315372495>
- 875 McElreath, R. (2020). *Rethinking: Statistical rethinking book package.*
- 876 Mettke-Hofmann, C., Lorentzen, S., Schlicht, E., Schneider, J., & Werner, F. (2009). Spatial neophilia and
 877 spatial neophobia in resident and migratory warblers (sylvia). *Ethology*, 115(5), 482–492.
- 878 Mikhalevich, I., Powell, R., & Logan, C. (2017). Is behavioural flexibility evidence of cognitive complexity?
 879 How evolution can inform comparative cognition. *Interface Focus*, 7(3), 20160121. <https://doi.org/10.1098/rsfs.2016.0121>
- 880 Morand-Ferron, J., Reichert, M. S., & Quinn, J. L. (2022). Cognitive flexibility in the wild: Individual dif-
 881 fferences in reversal learning are explained primarily by proactive interference, not by sampling strategies,
 882 in two passerine bird species. *Learning & Behavior*, 50(1), 153–166.
- 883 R Core Team. (2017). *R: A language and environment for statistical computing.* R Foundation for Statistical
 884 Computing. <https://www.R-project.org>
- 885 Revelle, W. (2017). *Psych: Procedures for psychological, psychometric, and personality research.* North-
 886 western University. <https://CRAN.R-project.org/package=psych>
- 887 Rutz, C., Klump, B. C., Komarzyk, L., Leighton, R., Kramer, J., Wischnewski, S., Sugarsawa, S., Morrissey,
 888 M. B., James, R., St Clair, J. J., et al. (2016). Discovery of species-wide tool use in the hawaiian crow.
 889 *Nature*, 537(7620), 403–407.
- 890 Seitz, K. A. M., Benjamin M. AND McCune. (2021). Using touchscreen equipped operant chambers to
 891 study animal cognition. Benefits, limitations, and advice. *PLOS ONE*, 16(2), 1–18. <https://doi.org/10.1371/journal.pone.0246446>
- 892 Selander, R. K., & Giller, D. R. (1961). Analysis of sympatry of great-tailed and boat-tailed grackles. *The
 893 Condor*, 63(1), 29–86.
- 894 Sol, D., Duncan, R. P., Blackburn, T. M., Cassey, P., & Lefebvre, L. (2005). Big brains, enhanced cognition,
 895 and response of birds to novel environments. *Proceedings of the National Academy of Sciences of the
 896 United States of America*, 102(15), 5460–5465. <https://doi.org/10.1073/pnas.0408145102>
- 897 Sol, D., & Lefebvre, L. (2000). Behavioural flexibility predicts invasion success in birds introduced to new
 898 zealand. *Oikos*, 90(3), 599–605. <https://doi.org/10.1034/j.1600-0706.2000.900317.x>
- 899 Sol, D., Székely, T., Liker, A., & Lefebvre, L. (2007). Big-brained birds survive better in nature. *Proceedings
 900 of the Royal Society of London B: Biological Sciences*, 274(1611), 763–769.
- 901 Sol, D., Timmermans, S., & Lefebvre, L. (2002). Behavioural flexibility and invasion success in birds. *Animal
 902 Behaviour*, 63(3), 495–502.
- 903 Stan Development Team. (2020). *RStan: The R interface to Stan.* <http://mc-stan.org/>
- 904 Stoffel, M. A., Nakagawa, S., & Schielzeth, H. (2017). rptR: Repeatability estimation and variance decom-
 905 position by generalized linear mixed-effects models. *Methods in Ecology and Evolution*, 8(11), 1639–1644.
- 906 Summers, J., Lukas, D., Logan, C., & Chen, N. (2023). The role of climate change and niche shifts in

- 910 divergent range dynamics of a sister-species pair. *Peer Community Journal*. <https://doi.org/10.24072/pejournal.248>
- 911 Taylor, A. H., Hunt, G. R., Holzhaider, J. C., & Gray, R. D. (2007). Spontaneous metatool use by new
912 caledonian crows. *Current Biology*, 17(17), 1504–1507.
- 913 Team, S. D. et al. (2019). Stan modeling language users guide and reference manual. *Technical Report*.
- 914 Tebbich, S., Sterelny, K., & Teschke, I. (2010). The tale of the finch: Adaptive radiation and behavioural
915 flexibility. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 365(1543),
916 1099–1109. <https://doi.org/10.1098/rstb.2009.0291>
- 917 Wehtje, W. (2003). The range expansion of the great-tailed grackle (*Quiscalus mexicanus gmelin*) in north
918 america since 1880. *Journal of Biogeography*, 30(10), 1593–1607. <https://doi.org/10.1046/j.1365-2699.2003.00970.x>
- 919 Wehtje, W. (2004). *The great-tailed grackle (Quiscalus mexicanus gmelin) in the western United States of
920 America: Range expansion and secondary contact between subspecies*. University of California, Riverside.
- 921 Wickham, H., François, R., Henry, L., & Müller, K. (2021). *Dplyr: A grammar of data manipulation*.
<https://CRAN.R-project.org/package=dplyr>
- 922 Wickham, H., Vaughan, D., & Girlich, M. (2023). *Tidyr: Tidy messy data*.
- 923 Wright, T. F., Eberhard, J. R., Hobson, E. A., Avery, M. L., & Russello, M. A. (2010). Behavioral flexibility
924 and species invasions: The adaptive flexibility hypothesis. *Ethology Ecology & Evolution*, 22(4), 393–404.
- 925 Xie, Y. (2013). Knitr: A general-purpose package for dynamic report generation in r. *R Package Version*,
926 1(7).
- 927 Xie, Y. (2017). *Dynamic documents with r and knitr*. Chapman; Hall/CRC.
- 928 Xie, Y. (2018). Knitr: A comprehensive tool for reproducible research in r. In *Implementing reproducible
929 research* (pp. 3–31). Chapman; Hall/CRC.
- 930
- 931
- 932