Peer Community Journal

Section: Ecology

Research article

Published
2025-08-18

Cite as
Kelsey B. McCune, Dieter Lukas, Maggie P. MacPherson and Corina J. Logan (2025) *Behavioral flexibility is related to exploration, but not boldness, persistence or motor diversity*, Peer Community Journal, 5: e83.

Correspondence
kelseybmccune@gmail.com

Peer-review
Peer reviewed and recommended by PCI Ecology,
https://doi.org/10.24072/pci.ecology.100771

This article is licensed under the Creative Commons Attribution 4.0 License.

Behavioral flexibility is related to exploration, but not boldness, persistence or motor diversity

Kelsey B. McCune[1,2], Dieter Lukas[3], Maggie P. MacPherson[1,4], and Corina J. Logan[3,5]

Volume **5** (2025), article e83

https://doi.org/10.24072/pcjournal.593

Abstract

Behavioral flexibility, the ability to change behavior when circumstances change based on learning from previous experience, is thought to play an important role in a species' ability to successfully adapt to new environments and expand its geographic range. However, behavioral flexibility is rarely directly tested at the individual level. This limits our ability to determine how it relates to other traits, such as exploration or persistence, that might also influence individual responses to novel circumstances. Without this information, we lack the power to predict which traits facilitate a species' ability to adapt behavior to new environments. We use great-tailed grackles (a bird species; hereafter "grackles") as a model to investigate this question because they have rapidly expanded their range into North America over the past 140 years. We evaluated whether grackle behavioral flexibility (measured as color reversal learning) correlated with individual differences in the exploration of new environments and novel objects, boldness towards known and novel threats, as well as persistence and motor diversity in accessing a novel food source. We determined that exploration of a novel environment across two time points and persistence when interacting with several different novel apparatuses were repeatable in individual grackles. There was no relationship between exploration or persistence and the two components of flexibility - the rate of learning to prefer a color option in the reversal learning task, and the rate of deviating from a preferred option. However, grackles that underwent serial reversal training to experimentally increase behavioral flexibility were more exploratory in that they spent more time in close proximity to the novel environment relative to control individuals. This indicates that, the more an individual investigated a novel apparatus, the more it was able to potentially learn and update its knowledge of current reward contingencies to adapt behavior accordingly. Our findings improve our understanding of the traits that are linked with flexibility in a highly adaptable species. We highlight the importance of using multiple different methods for measuring boldness and exploration to evaluate consistency of performance and therefore the methodological validity. We also show a link between exploration and behavioral flexibility that could facilitate adaptation to novel environmental changes.

[1]Institute for Social, Behavioral and Economic Research, University of California Santa Barbara, USA, [2]College of Forestry, Wildlife and Environment, Auburn University, USA, [3]Department of Human Behavior, Ecology and Culture, Max Planck Institute for Evolutionary Anthropology, Germany, [4]Department of Biology, Trent University, Canada, [5]Neuroscience Research Institute, University of California Santa Barbara, USA

Peer Community Journal is a member of the Centre Mersenne for Open Scientific Publishing
http://www.centre-mersenne.org/

e-ISSN 2804-3871

Introduction

Humans are altering all ecosystems on the planet too rapidly for most species to evolve adaptations to survive and reproduce (Hendry et al., 2008; Sih, 2013). Among other consequences, anthropogenic change can lead to a proliferation of novel habitats, foods, and predators (Sih et al., 2011). Across short timescales, individuals must adapt to this novelty through changes in behavior. Behavioral flexibility (hereafter “flexibility”) is defined as the ability to use learning to functionally change behavior when circumstances change (Mikhalevich et al., 2017). As such, flexibility is thought to facilitate species resilience to anthropogenic change (Sol et al., 2013) and species invasions into novel areas (Sol et al., 2002; Wright et al., 2010).

The relationship between flexibility and adaptation to anthropogenic change is rarely directly tested at the individual level. Research studying the impact of flexibility on the success of species invasions most often uses proxies of flexibility such as species’ brain size, or presence of the theoretical outcomes of flexible behavior like the number of foraging innovations (Sol et al., 2002). The few studies that have directly related environmental adaptation to flexibility through measures of reversal learning show that flexible behavior can be closely linked with the current environmental niche. For example, mountain chickadees that live in harsh, high elevation environments perform worse on reversal learning tasks relative to lower elevation, milder climate individuals (Croston et al., 2017). This suggests that individuals that have a wider range of food options and a reduced reliance on cached food in milder climates require more flexibility to switch between food types. Additionally, new evidence from great-tailed grackles shows that the more flexible individuals also demonstrate greater foraging diversity in the wild (Logan et al., 2025), and were better able to innovate solutions on a novel foraging apparatus (Logan et al., 2023). Consequently, flexibility may show variation within, as well as among, species and may affect diverse aspects of individual behavioral interactions with the environment. To better understand how flexibility might facilitate responses to novelty and resilience to anthropogenic change, it is important to directly test flexibility and relate it to other ecological and behavioral traits at the individual level.

Although behavioral flexibility has been the trait that much research has focused on to understand how behavior can impact adaptation to anthropogenic environmental changes, individual differences in other traits like exploratory tendency, boldness, persistence, or motor diversity could also play a role and correlate with behavioral flexibility (Sol et al., 2002; Logan, 2016a). To distinguish whether observed behavior in the wild or performance on behavioral trait assays are motivated by one or more distinct traits, it is important to measure multiple traits in the same individuals (Carter et al., 2013). However, evaluation of the relationship between flexibility and other behavioral traits has produced inconsistent results (Logan, 2016a; Dougherty & Guillette, 2018). In one well studied avian group, the Paridae, flexibility is related to exploration, which increases the likelihood of encountering fitness-enhancing resources in novel environments (Canestrelli et al., 2016; Griffin et al., 2016). This might imply that they are not two distinct traits, but the direction of the relationship is inconsistent across species (negative: Amy et al., 2012; positive: Rojas-Ferrer et al., 2020). Individuals approaching a potentially threatening aspect of the environment require a certain degree of boldness (McCune et al., 2018). However, the relationship between boldness and flexibility can be positive (Titulaer et al., 2012), negative (Bebus et al., 2016; Bensky & Bell, 2022), or neutral (Guenther et al., 2014; De Meester et al., 2022). Theoretically, persistence should inhibit flexibility because it results in perseverating on a previously rewarded behavior rather than changing to a more productive behavior for a given circumstance (Morand-Ferron et al., 2022). In contrast to persistence, motor diversity is theoretically positively correlated with flexibility because it implies that the individual has a repertoire of different behaviors it is able to choose from to match each circumstance (Diquelou et al., 2016). Research in squirrels supports this prediction (Chow et al., 2016), where the more flexible individuals were less persistent and more likely to use diverse motor behaviors. Whereas, an earlier study in great-tailed grackles using different behavioral assays found no relationship between flexibility and any other behavioral traits, including persistence and motor diversity (Logan, 2016a).

The lack of consistent support for which behavioral traits are related (or not) to flexibility could stem from what has been called a "jingle-jangle fallacy" (Carter et al., 2013). This term describes the mismatch between a trait label (like exploration) and what the method (novel environment) actually measures (could be exploration, activity, or boldness). A mismatch can occur when researchers use a single trait label for what are actually multiple distinct inherent traits ("jingle fallacy"), or if using two or more distinct labels for what is actually the same inherent trait ("jangle fallacy"). One step towards avoiding this issue is to use multiple experimental methods, as in a test battery, to measure a variety of behaviors, then assess the relationships among performance to identify which aspects of the behaviors that are measured might be driven by the same underlying trait (Perals et al., 2017; Shaw & Schmelz, 2017).

To determine whether behavior labels represent the same underlying trait, it is also important to ensure that measured performance on behavioral assays is consistent within individuals across time and context (i.e., repeatable). Inter-individual differences in performance could result from short-term variation in the external environment like social interactions or food availability or variation in internal states like hunger or stress. This plasticity is distinct from consistent individual differences in behavior across contexts stemming from genetic or developmental effects (i.e., animal personality; Fidler et al., 2007; Duckworth, 2010). If behavioral traits are heritable, multiple traits can become linked through natural selection such that individuals that show high values on one trait (e.g., flexibility), will consistently display high values on a linked trait (e.g., exploration; Reale et al., 2007; Rowe & Healy, 2014). It is important to know whether traits are linked because such linkage could result in limited behavioral plasticity that may alter the ability or mode of adapting to rapid environmental changes (Sih et al., 2004). Indeed, inconsistency in the direction of the relationship between flexibility and behavioral traits in previous studies could stem from a lack of repeatability in performance on behavioral trait assays. To address whether flexibility is related to other behavioral traits, we must first assess whether our methods produce performance that is repeatable (Dingemanse & Dochtermann, 2013) to validate that it is more likely to represent variation in a heritable trait.

In a previous study with a smaller sample size (Logan, 2016a), we found no evidence for significant correlations between flexibility and the behavioral traits exploration, boldness, persistence, and motor diversity. However, this result could stem from the small sample size and lack of power to detect a relationship with a small effect size, or methods that do not result in repeatable performance. Based on this preliminary evidence, in the present study we increased our power to detect a relationship by training some individuals to be more flexible before measuring the other behavioral traits. Additionally, we tested whether performance on measures of exploration, boldness, persistence, and motor diversity are repeatable across time and contexts and therefore likely represent distinct personality traits. Behavior is considered repeatable if the variance in performance on the task is smaller within individuals compared to the variance among individuals. If there is no repeatability of these behaviors within individuals, then performance is likely state dependent (i.e., it depends on fluctuating motivation, stress, hunger levels, etc.) and/or is reliant on the current context of the tasks, and therefore less likely to consistently correlate with flexibility (Griffin et al., 2015). Then, we assessed whether the repeatable traits were related to performance on a flexibility task. We focus on great-tailed grackles (*Quiscalus mexicanus*; hereafter "grackles") because they are likely to have experienced selection for behavioral adaptations to rapid environmental change. Grackles have rapidly expanded their range into novel areas in North America over the past 140 years (Wehtje, 2003; Summers et al., 2023) and our previous research on this species has demonstrated that grackles are flexible (Logan, 2016b), and that flexibility is a distinct trait on which grackles show individual variation (Logan et al., 2023). Thus, this species is ideal for assessing whether flexibility is part of a suite of behaviors that facilitate adaptation to novel environments.

Preregistered hypotheses and predictions summary

We preregistered several additional predictions pertaining to alternative measures of behavioral flexibility, however we include below only those that ended up being relevant. See the preregistration for the full list and the criteria that determined which variables to use (Supplementary Material 3). This article also attempt to test a Hypothesis 2 and its associated predictions, which are reported in Supplementary Material 2. The prediction numbers listed here maintain the original order from the preregistration to help readers track consistency across Stage 1 and Stage 2.

Hypothesis 1: Behavioral flexibility is correlated with the exploration of new environments and novel objects, but not with boldness, persistence, or motor diversity.

Predictions 1-5: Behaviorally flexible individuals will be more exploratory of novel environments (P1) and novel objects (P2) than less flexible individuals, but there will be no difference in persistence (P3), boldness (P4), or motor diversity (P5) (as found in Logan, 2016b).

P1 alternative 4: There is no correlation between exploration and behavioral flexibility because our novel object and novel environment methods are inappropriate for measuring exploratory tendency. These measures of exploration both incorporate novelty and thus may measure boldness rather than exploration. This will be supported by a positive correlation between behavioral responses to our exploration and boldness assays.

P3 alternative 1: There is a positive correlation between persistence and the number of incorrect choices in reversal learning before making the first correct choice. This indicates that individuals that are persistent in one context are also persistent in another context.

P3 alternative 2: There is no correlation between persistence and the number of incorrect choices in reversal learning before making the first correct choice. This indicates that flexibility is an independent trait.

Methods

Preregistration details

The hypotheses, methods, and analysis plan are described in detail in the peer-reviewed preregistration, in Supplementary Material 3. We summarize these methods here, with any changes from the preregistration noted in the *Changes after the study began* section. The preregistration was written and submitted to Peer Community In (PCI) Ecology for peer review (Sep 2018) before collecting any data. After data collection began (and before any data analysis was conducted), we received peer reviews from PCI Ecology, revised, and resubmitted the preregistration (Feb 2019). It received an in principle recommendation in Mar 2019.

Subjects

Grackles were caught in the wild in Tempe, Arizona USA using mist nets, walk-in traps and bow nets. Trapping could occur at any time of day when grackles were active. While some trapping methods can select for subjects with certain traits (e.g., boldness Biro & Dingemanse, 2008; but see Brehm & Mortelliti, 2018), mist nets are not visible to birds and no habituation is required, decreasing the probability of a selection bias for individuals that are more bold, food motivated, etc. Grackles were then individually housed in an aviary (each 244 cm long by 122 cm wide by 213 cm tall). We aimed for a balanced sample of adult males and females, but because grackles in this population were difficult to catch, we ultimately ended up with only 4 females (15 males) and 2 juveniles (17 adults). Grackles were held in captivity until they completed the test battery, or 6 months had passed. All grackles were then released back into the wild and subsequently observed exhibiting normal behavior.

Test battery

During testing (except exploration, see below) we food deprived grackles for up to four hours per day, but they had the opportunity to receive high value food items by participating in the assays. They had access to a maintenance diet at all other times, and access to water at all times. Individuals were given three to four days to habituate to the aviaries before their test battery began. Birds were then tested 6 days per week. On each testing day, we conducted multiple testing sessions where the duration of the session depended on the grackle's motivation to participate or the task design (see below).

We use data from a recent investigation (Logan et al., 2023; Lukas et al., 2024) on the flexibility of 19 grackles, and here we additionally measured exploration and boldness in these same individuals. We also measured persistence and motor diversity through performance on two multiaccess boxes (MABs) in 17 of these grackles. The research described here is part of a larger project where the main goal was to better understand the impact of flexibility on diverse cognitive, behavioral, and physiological traits. Consequently, for all grackles we first assayed flexibility and implemented a flexibility training where half of the grackles underwent serial reversal learning and the other half received only one reversal and then control trials, described below. The training resulted in grackles more quickly changing their behavior when reward contingencies changed, relative to control grackles (Logan et al., 2023). By experimentally increasing the difference in behavioral flexibility between control and trained grackles, we increased our power to detect relationships between flexibility and other traits. After grackles passed the flexibility training, they received the subsequent behavioral trait assays in a randomized order. Grackles were assayed twice for exploration and boldness, and given sessions with the MABs until they passed criterion. Because there were two MABs, we also have two measures of persistence and motor diversity for each individual.

Behavioral flexibility

We used the reversal learning paradigm to measure flexibility as the ability to change behavior when circumstances change. In the first phase of reversal learning, subjects learn an initial association between a stimulus (here, color) and food. The reversal phase then occurs where the food is switched to the other color and the measure of flexibility is how quickly the subject learns the new food-color association. The methods for the initial association and the reversal trials are identical, where, on each trial, grackles could choose to look inside one of two colored containers for food (Figure 1a). After they make a choice, the experimenter removes both containers, refills the food if necessary, then replaces the containers for the next trial. The side that the rewarded container was on was pseudorandomized to never be on the same side more than twice in a row to inhibit grackles from forming a side bias. When grackles showed a significant preference for the rewarded color in the initial association phase, demonstrated by choosing correctly on 17 out of the most recent 20 trials, we switched the location of the food to the other colored container (a "reversal"). We measured baseline flexibility as the number of trials it took grackles to choose correctly on 17 out of the most recent 20 trials in this first reversal to demonstrate a change in preference to the second colored container. The flexibility training consisted of a randomized subset of grackles (n = 8) that received serial reversals where we switched the location of the food in multiple reversals after the grackle passed criterion in each reversal. Serial reversals continued until grackles were switching their preference in each reversal quickly enough to meet our experiment's passing criterion of two consecutive reversals in 50 trials or fewer. We chose a criterion of 50 trials based on an earlier study of grackle reversal learning performance (Logan, 2016a) where 50 represented an approximately 30% increase in the speed that grackles switched their preference in the first reversal (Logan et al., 2023). Grackles needed 6-8 reversals to pass this serial reversal training. Instead of serial reversals, control grackles (n = 11) received equal testing experience with two identically colored containers, both containing a food item.

From the performance of each individual on reversal learning, we used Bayesian reinforcement learning models to create the Flexibility Comprehensive variables by modeling all of the choices that individuals made during the serial reversal learning experiment, and the uncertainty around these choices. Because we include the sequence of all correct and incorrect choices individuals made during reversal learning, these variables more effectively represent flexibility compared to more commonly used variables such as the number of trials to reverse a preference. The details of this model and the validation of it as a measure of flexibility are described elsewhere (Blaisdell et al., 2021; Lukas et al., 2024). The Flexibility Comprehensive variables consist of two components: ϕ (the Greek letter phi) as the rate of learning to be attracted to a color option and λ

(the Greek letter lambda) as the rate of deviating from learned attractions that were previously rewarded. Thus, our two measures of flexibility, that we subsequently included as covariates explaining behavioral trait performance, were the Flexibility Comprehensive continuous variables or the dichotomous variable describing whether the grackle was in the flexibility trained or control group. There was one measure per individual for each of these variables.

All measures of the behavioral traits exploration, boldness, motor diversity, and persistence were collected after the serial reversal learning training was complete. By experimentally increasing the difference in flexibility performance between the trained and control grackles we increased our ability to detect a relationship, if it exists, between this trait and the other traits under investigation in this study.

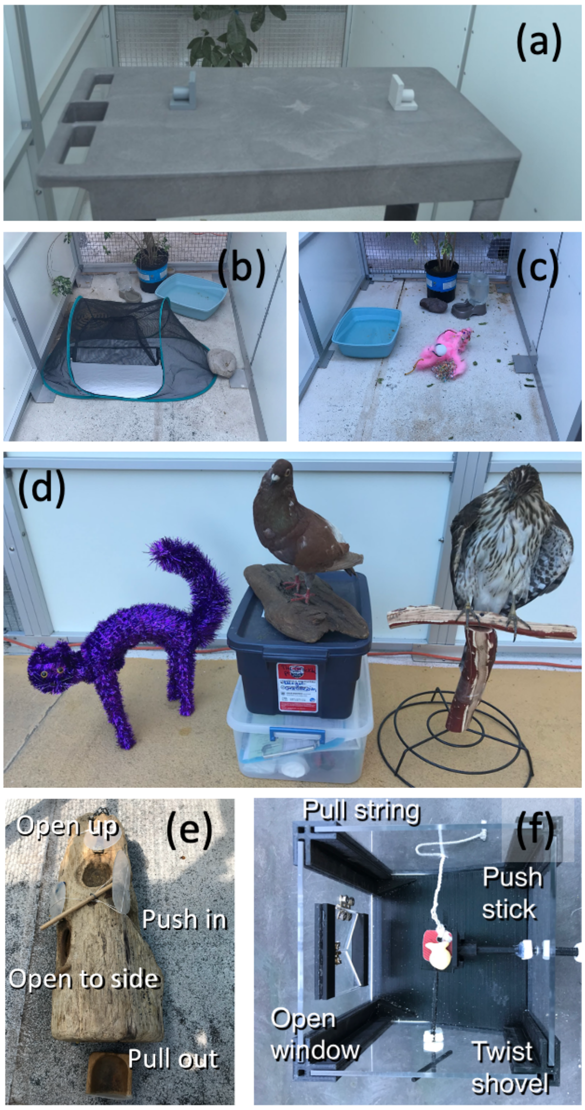

Figure 1 - This experiment assessed the relationship between multiple different behavioral tests and contexts. We quantified and increased behavioral flexibility with serial reversal learning of a color preference: a light gray and a dark gray tube (a), we determined individual differences in exploration of a novel environment: a tent (b), exploration of a novel object: a homemade pink fuzzy shape (c), boldness towards threatening objects (purple halloween cat and Cooper's hawk) compared to a known non-threat (pigeon) (d), we cataloged motor diversity when interacting with novel foraging problems on the two multiaccess boxes (e-f), and we measured persistence by counting the number of touches to all novel apparatuses (b, c, d, e, and f).

Behavioral traits

Boldness

We define boldness as an individual's response to a potential threat (Reale et al., 2007). We measured boldness with two different threatening objects, a known threat (taxidermied Cooper's hawk) and a novel threat (purple cat halloween decoration). We also included a known non-threat (taxidermied pigeon) as a control condition (Figure 1d). Each individual was assayed with all three objects, presented in randomized order, across three days. Exposure to each object was limited to 15 minute trials, and a food item was placed next to the object. Boldness assays occurred while the grackle was food deprived to elicit approach behaviors. We conducted each of these assays twice to measure the repeatability of performance on this task to verify that the experimental designs elicited behaviors indicative of an inherent personality trait (as opposed to a passing motivational state). During boldness trials we measured multiple behaviors and, as preregistered, statistically analyzed the variable for which we ultimately had the most data, "Duration on the Ground", encompassing the total time grackles spent within 100 cm of the object.

Exploration

We defined exploration as an individual's response to novelty (Reale et al., 2007) to gather information that does not satisfy immediate needs (Mettke-Hofmann et al., 2002). We used two different assays to measure exploratory tendency: novel environment (a small tent) and novel object (a pink fuzzy shape) exploration (Figure 1b & 1c). We also conducted control conditions where we measured the grackle's behavior in its familiar environment (the aviary) and with a familiar object (an empty water dish). Exploration tests occurred when the grackle was not food deprived to ensure that any approach to the novel object was for information gathering rather than food. Each trial was 45 minutes long and we always conducted the familiar condition trial immediately before the novel condition trial. We also conducted each of these assays twice to measure repeatability. As in boldness trials, we measured multiple behaviors during exploration trials and statistically analyzed the variable for which we ultimately had the most data. In the exploration of the novel environment condition we had the most data for two variables, "Duration near" (within 20cm) and "Latency to first land on the ground" within 100 cm of the object, so we conducted one model for each variable. For the exploration of the novel object condition, we had the most data for "Latency to first land on the ground".

Motor diversity and persistence

We defined motor diversity as the number of different motor actions used to solve novel problems on either of two multiaccess boxes (MABs; Figure 1e & 1f). We used an ethogram (Table 1) to define and distinguish each interaction with the MABs. For each grackle, we summed the number of distinct motor actions they used while interacting with each MAB, resulting in two values for each grackle. We quantified persistence as the number of touches to a novel apparatus per trial time (Griffin et al., 2015; Logan, 2016a), where the novel apparatuses included the novel environment and novel object from the exploration assays, the potentially threatening boldness objects, as well as the two MABs. We summed the number of touches grackles made to each apparatus, resulting in a value of persistence for each test apparatus, if the grackle received that test (e.g., two grackles did not participate in the MAB tests). We further distinguished touches to the MABs based on whether they were functional (touches to the doors or loci that could result in getting the food item) or nonfunctional (touches to the side of the box that would never result in food). Motor diversity and persistence were coded from videos of grackles interacting with the two different MAB apparatuses for a separate experiment on problem solving ability (Logan et al., 2023), as well as the novel apparatuses from the exploration and boldness assays.

Table 1 - Motor action ethogram for the two multiaccess box experiments. Any of the four modifiers can be added to any of the six motor actions. However, Stand only goes with the On top modifier, resulting in a total of 21 unique motor actions. For example, Vertical Peck is a peck to a vertical surface, and Gape Upside Down is a gape with the head being held upside down. Note that one interaction can be coded in multiple categories (e.g., if a bird pulls the string first horizontally and then vertically).

Body part	Motor action	Description
Bill	Peck	Touches the apparatus or its pieces, usually a short duration (e.g., 1s). A peck is with the bill closed or open, but just the tip of the bill touches the apparatus.
	Push	Moves away from self a piece of the apparatus or its pieces, usually of a longer duration than a peck.
	Pull	Moves towards self a piece of the apparatus or its pieces, usuallly of a longer duration than a peck
	Grab	Holds a piece of the apparatus or its pieces, usually of a longer duration than a peck. The bill will be open and the part of the bill touching the apparatus will be the inside of the mandibles.
	Gape	The closed bill is placed under the edge, in an opening, or on a surface of the apparatus or its pieces and the the bill is opened. Usually of a longer duration than a peck.
Feet	Stand	Oriented upright on top of the apparatus.
	Modifiers	**These can apply to any of the above actions**
	Vertical (e.g., head vertical to the ground)	Performs an action directed vertically, often towards the horizontal (oriented parallel to the ground) edges of the apparatus (e.g., the lid of the box), or moves a piece of the apparatus up.
	Horizontal (e.g., head parallel to the ground)	Performs an action directed horizontally, often towards the vertical (oriented upright to the ground) edges of the apparatus (e.g., the walls), or moves a piece of the apparatus horizontally.
	Upside down	Performs an action with its head upside down.
	On top	Performs an action while standing on top of the apparatus.

Statistical analyses

General analysis plan

For all analyses, we used the MCMCglmm function in the MCMCglmm R package (Hadfield, 2010). Our preregistered analysis plan was to use a Poisson distribution and log link for both the repeatability analyses and analyses testing the correlation of behavioral traits with flexibility. However, we used the DHARMa package (Hartig, 2022) to verify that the data for each analysis met the assumptions for Poisson regression and modified the model family accordingly (see below in *Changes after the study began*). We started each model with 13,000 iterations, a thinning interval of 10, a burnin of 3,000, and minimal priors (V=1, nu=0). We checked that the GLMM showed acceptable convergence (i.e., lag time autocorrelation values <0.01 Hadfield, 2010), and adjusted the number of iterations, thinning and burnin if necessary. Due to our unbalanced sample of sex and age we checked whether these variables significantly impacted the response. We found that these covariates did not have a significant effect on any of the models (described below), so we omitted them from the final models (see *Changes after the study began* section).

Repeatability

We obtained repeatability estimates that account for the observed and latent scales. The repeatability estimate indicates how much of the total variance, after accounting for fixed and random effects, is explained by individual differences. For each behavioral trait, we included fixed effects to control for variation in the response not attributable to individual differences and consequently we report the adjusted repeatability estimates. All models included a covariate describing whether the grackle was flexibility trained or in the control group. Our boldness model additionally included a covariate for threat condition (hawk or cat). For persistence, we additionally included a covariate for assay type and one for the total time the grackle had access to an assay to control for opportunity to make functional or non-functional touches. The motor diversity model included an additional covariate for assay type. Marginal and conditional R-squared values are reported in Table S1 of Supplementary Material 2 to illustrate the impact of fixed effects on repeatability estimates.

From the posterior distribution of the MCMCglmm model for each behavioral trait, we extracted the Bird ID random effect variance to calculate the ratio of variance accounted for by individual differences relative to total variance (Nakagawa & Schielzeth, 2010). We used the mean value of this ratio across all iterations for a given behavioral trait as our measure of repeatability. We used the HPDinterval function from the coda package (Plummer et al., 2024) to calculate credible intervals around our repeatability estimate. We used permutation tests that randomized data among individuals to test the significance of the repeatability value.

Relationship with flexibility

If performance was repeatable across two time points in the behavioral trait assays, we used the average value per bird per assay in Bayesian multivariate models to investigate whether performance was related to the Flexibility Comprehensive variables (ϕ and λ). As such, the performance variables from each behavioral trait assay were the dependent variables and ϕ and λ were the independent variables. We assessed the relationship between flexibility and the behavioral trait by interpreting the parameter estimates from these models. Similarly, we used Bayesian bivariate models to analyze whether there was a difference in performance on the behavioral trait assays between grackles that underwent serial reversal learning flexibility training relative to grackles in the control group.

Changes after the study began

After data collection began and before data analysis

1) We added an *unregistered analysis* to assess interobserver reliability for the response variables to determine how repeatable our data collection was by having the videos coded by multiple coders. This unregistered analysis is described, and results reported, in the Supplementary Material 1.

After data collection and during data analysis

1) We conducted an *unregistered analysis* to compare the grackles' responses to the familiar item with responses to the novel/threatening items in the exploration and boldness assays. The definition for boldness relates to the behavioral response to threat, so we would expect a decrease in interactions with the novel/threatening items relative to the control item. To test that this occurred, and the grackles perceived the items as threatening, we used MCMCglmm to model the effect of condition (novel or familiar item trial) on the latency to approach and the duration spent in proximity to the items in the exploration assays. We used a gaussian distribution for latency to approach and Poisson distribution for the duration spent in proximity. We included a covariate that identified whether the bird was in the flexibility trained (or control group) and a random effect for bird ID. The boldness data were overdispersed and zero-inflated so we used a zero-inflated negative binomial mixed model with the R package NBZIMM (Zhang & Yi, 2020). In this model, we also included a covariate for the flexibility trained group and a random effect for bird ID.

2) For the repeatability analyses, we preregistered that we would calculate repeatability from the ratio of variance components extracted from MCMCglmm models. We also obtained credible intervals from the posterior distribution of these models. However, repeatability is a ratio so values can never be less than zero. As such, we are not able to ascertain the significance of our repeatability values by determining whether the credible interval overlaps with zero. We conducted an *unregistered analysis* to obtain p-values indicating whether performance was significantly more repeatable than random by utilizing the built in permutation tests in the rptR package (Stoffel et al., 2017). This also ensured that repeatability values and credible intervals were consistent with the preregistered MCMCglmm methods to validate that our non-informative priors were appropriate.

3) The boldness data were zero-inflated (69% of the data were zeros) and overdispersed, such that the appropriate model for this kind of count data is a zero-inflated negative binomial model. As stated above, we used this model type in the *unregistered analysis* to compare the responses between the threatening and non-threatening contexts. To assess repeatability of performance on the boldness assays, we preregistered that we would use a MCMCglmm model with a Poisson

distribution. The boldness data were not appropriate for Poisson and we do not know of a method for obtaining the variance components for the repeatability calculation from a zero-inflated negative binomial model. Consequently, for the repeatability analysis we used a logistic regression, where the response was 0 (the grackle never approached the object during boldness trials) or 1 (the grackle approached the object during boldness trials).

4) For repeatability analyses of the exploration and persistence data, we originally planned to conduct a model with a Poisson distribution. However, the data checking process detected significant zero-inflation and heteroscedasticity in the Poisson models. We log-transformed the latency to approach (for exploration) and number of touches (for persistence) for the gaussian model, which was normally distributed and not heteroscedastic, therefore we used a gaussian distribution instead.

5) When we originally submitted this preregistration, we anticipated measuring motor diversity on only one multiaccess box (MAB). However, as part of a different experiment within our overall project, we added a second, but distinct MAB. Consequently, we did not preregister a repeatability analysis for motor diversity because there would have been only one measure per bird. We added an *unregistered analysis* to assess motor diversity repeatability. We used a Poisson regression and included a covariate for whether the grackle was flexibility trained or not. We also included an offset for the total trial time with the MABs to control for variation in the opportunity to express motor behaviors.

6) During the exploration environment assays, very few grackles stepped inside the tent (n = 4), so we did not have enough data to use the following preregistered variables in the analysis relating exploration and behavioral flexibility: Latency to enter a novel environment inside a familiar environment, Time spent in each of the different sections inside a novel environment or the corresponding areas on the floor when the novel environment is not present (familiar environment) as an interaction with the Environment Condition: activity in novel environment vs. activity in familiar environment, Time spent per section of a novel environment or in the corresponding areas on the floor when the novel environment is not present (familiar environment) as an interaction with the Environment Condition: time spent in novel environment vs. time spent in familiar environment.

7) We also realized that, because we experimentally increased reversal learning speed through serial reversal learning (Logan et al., 2023), behavioral flexibility should be the independent rather than dependent variable.

8) We found (Blaisdell et al., 2021; Lukas et al., 2024) that the “Flexibility Comprehensive” variables were much more effective at representing flexibility than the other variables we preregistered (e.g., Trials to reverse in the last reversal). Additionally, we found that solution switching on the MAB is correlated with reversal performance and including this as an additional variable describing flexibility will not significantly add to the variance explained. Because the individual’s serial reversal learning training condition (control or trained) is accounted for in the Flexibility Comprehensive variable, we did not include condition as an additional independent variable in these models. Note that we still conducted the preregistered analyses testing the relationship between performance on the behavioral trait assays and whether the individual was in the control or flexibility trained group.

9) We preregistered that we would include “Age” as a covariate in our models relating performance on the behavioral trait assays to flexibility if we tested juveniles as well as adults, though our plan was to only test adults. Our sample ultimately included two juveniles because the grackles were more difficult to catch than expected and we struggled to meet our minimum sample size. Similarly, it is possible that Sex could influence performance, but we only tested 4 females because they were more difficult to trap than males. We did not find that including a covariate for Age and Sex changed any of our results (repeatability or relationship with flexibility). Therefore, to maintain greater statistical power, we decided not to include Age or Sex as covariates in the final models.

10) We added an additional persistence repeatability analysis to test whether nonfunctional touches were consistent across the two different MABs. We preregistered that we would separately

evaluate the relationship between flexibility and functional or nonfunctional touches, but, because flexibility was originally the dependent variable, we did not preregister this repeatability analysis.

11) We made two modifications to the analysis testing the relationship between persistence and flexibility. We preregistered that we would use all of the data, including the repeated measures, with a random effect for individual ID in a Poisson model. However, the full data set was zero-inflated. Because persistence was repeatable across assays, we took the average for each individual to use as the dependent variables in our models. Consequently, there was no potential for within-individual clustering in the data and we did not include the random effect for individual ID. Secondly, we were interested in the number of touches to novel objects per time. As such, we used a Poisson model as preregistered, but with an added offset term for trial time.

12) We preregistered that we would compare performance on the boldness and exploration assays between grackles in the aviaries and those tested in the wild. However, we were unable to collect a large enough sample size to quantitatively test this hypothesis, therefore we present what we have in Supplementary Material 2.

Results

Repeatability

Our first goal was to assess the repeatability of grackle boldness, exploration, persistence and motor diversity behaviors across time and different contexts. We collected boldness and exploration data on 19 individuals, but 2 of these individuals did not participate in the MAB tasks and so our sample size was 17 for the repeatability of persistence and motor diversity.

Boldness

We first conducted an *unregistered analysis* to evaluate whether grackles perceived the objects presented to them during boldness trials as threatening. Relative to the pigeon control condition (the known non-threat), we found that grackles spent 55% less time on the ground within 100 cm of the cat ($p = 0.00$) and 61% less time on the ground in the presence of the hawk ($p = 0.00$). There was a nonsignificant 9.5% decrease in duration on the ground in the hawk condition relative to the cat condition ($p = 0.71$). Consequently, there is evidence that the grackles perceived the cat and hawk as more threatening than the pigeon, and we only use data from the cat and hawk assays in all subsequent analyses including boldness. Despite the perceived threat, 12 out of 19 grackles spent time on the ground in the presence of the hawk and 7 out of 19 grackles spent time on the ground with the cat at some point during the 15-minute boldness trials.

Next, we assessed whether grackles reacted consistently towards each threatening object across two time periods (temporal repeatability). Because the repeatability analysis was not possible with a zero-inflated negative binomial model, we instead used a binomial model where our dependent variable represented whether the duration grackles spent within 100 cm of the threatening object was greater than 0 seconds (1) or not (0). We found no evidence for repeatability of performance in either the cat (Repeatability = 0.18, CI = 0.00-0.96, $p = 0.22$) or hawk (R = 0.00, CI = 0.00-0.44, $p = 0.48$) assays (Figure 2). Similarly, when we considered grackle performance across the two different threatening contexts (contextual repeatability) there was also no consistency in behavioral response (R = 0.04, CI = 0.00-0.28, $p = 0.22$).

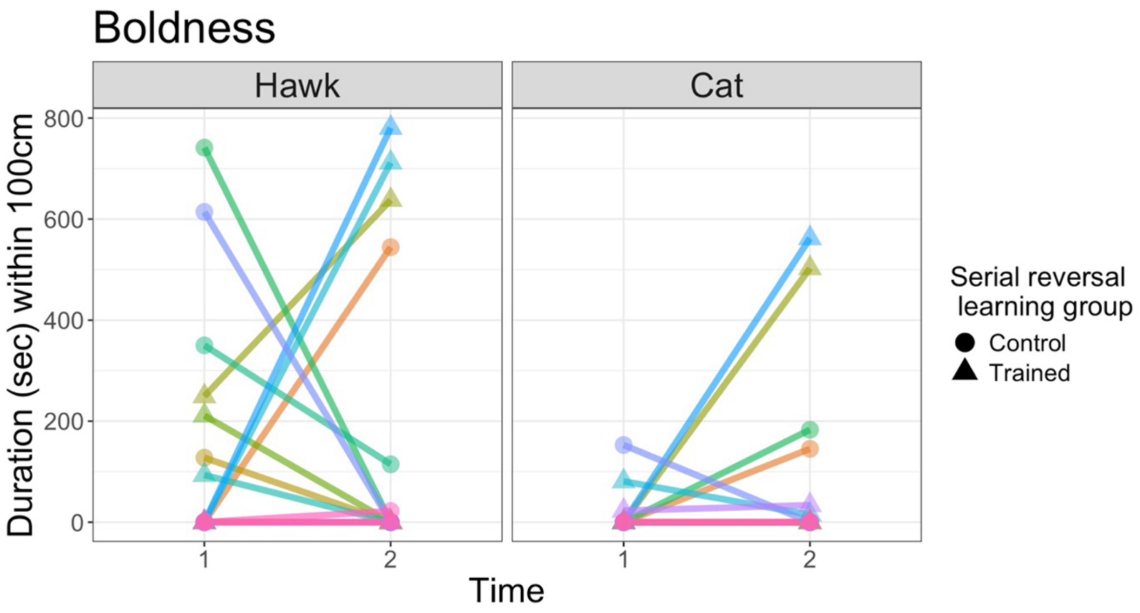

Figure 2 - The grackles did not respond consistently to the threatening objects across the two time points. Each line color represents an individual and the points show the number of seconds individuals spent on the ground within 100 cm of the threatening object during each of the two 15-minute trials (Time 1 and Time 2). The shape of the point is based on whether the grackle was part of the control (circle) or trained (triangle) group in the serial reversal learning experiment. The two time periods were separated by an average of 33 days (range: 11-49 days). If performance is repeatable we would expect the line connecting the two points to be at or close to horizontal, and the lines of different individuals to be approximately parallel.

It is possible that the lack of repeatability is because habituation to the potentially threatening object occurs after the first exposure (Greggor et al., 2015; Takola et al., 2021). We conducted an *unregistered analysis* and found that grackles did spend significantly longer on the ground during the second cat, hawk and novel object (which the grackles considered threatening, see below) trials, relative to the first trials (Poisson model: $ß = 0.85$, $p < 0.01$). To check whether this explains the lack of contextual repeatability in this behavioral trait, we conducted a second *unregistered analysis* evaluating repeatability of performance in only the first trial in response to the potentially threatening contexts. We still found no evidence that response to the potentially threatening objects was repeatable across these contexts (R = 0.00, CI = 0.00-0.17, $p = 1$; Figure S3).

Exploration

Similar to boldness, we assessed the repeatability of exploratory behavior across two time points and across two different contexts: a novel object and a novel environment. Because novel items might elicit a response based on the boldness personality trait rather than an exploratory response (our P1 alternative 4 described above; Carter et al., 2013), we also compared the novel environment and novel object responses to control conditions with a familiar environment and a familiar object to determine whether grackles perceived the novelty as threatening (note that this is an *unregistered analysis*). We found no difference in the latency of individuals to approach the novel compared to the familiar environment ($ß = 0.29$, CI = -0.24-0.81, $p = 0.27$), or the duration they spent near the novel and familiar environments ($ß = -0.61$, CI = -1.47-0.20, $p = 0.14$). In contrast, grackles took significantly longer to approach the novel object relative to the familiar object ($ß = 2.11$, CI = 1.22-2.89, $p < 0.01$), indicating the novel object may have been perceived as threatening.

We found that the latency to approach the novel environment across time points 1 and 2 was highly repeatable (R = 0.72, CI = 0.42-0.88, $p < 0.01$). Similarly, the duration spent near the novel environment was also highly repeatable (R = 0.85, CI = 0.67-0.98, $p < 0.01$). However, the latency to approach the novel object was not repeatable (R = 0.05, CI = 0-0.5, $p = 1$; Figure 3). When we assessed performance across the novel environment and novel object tasks, we found that latency to approach was repeatable across the two different contexts, but this result was driven by the very high between-individual variance in the environment assay (R = 0.49, CI = 0.21-0.69, $p = 0$; Figure S1).

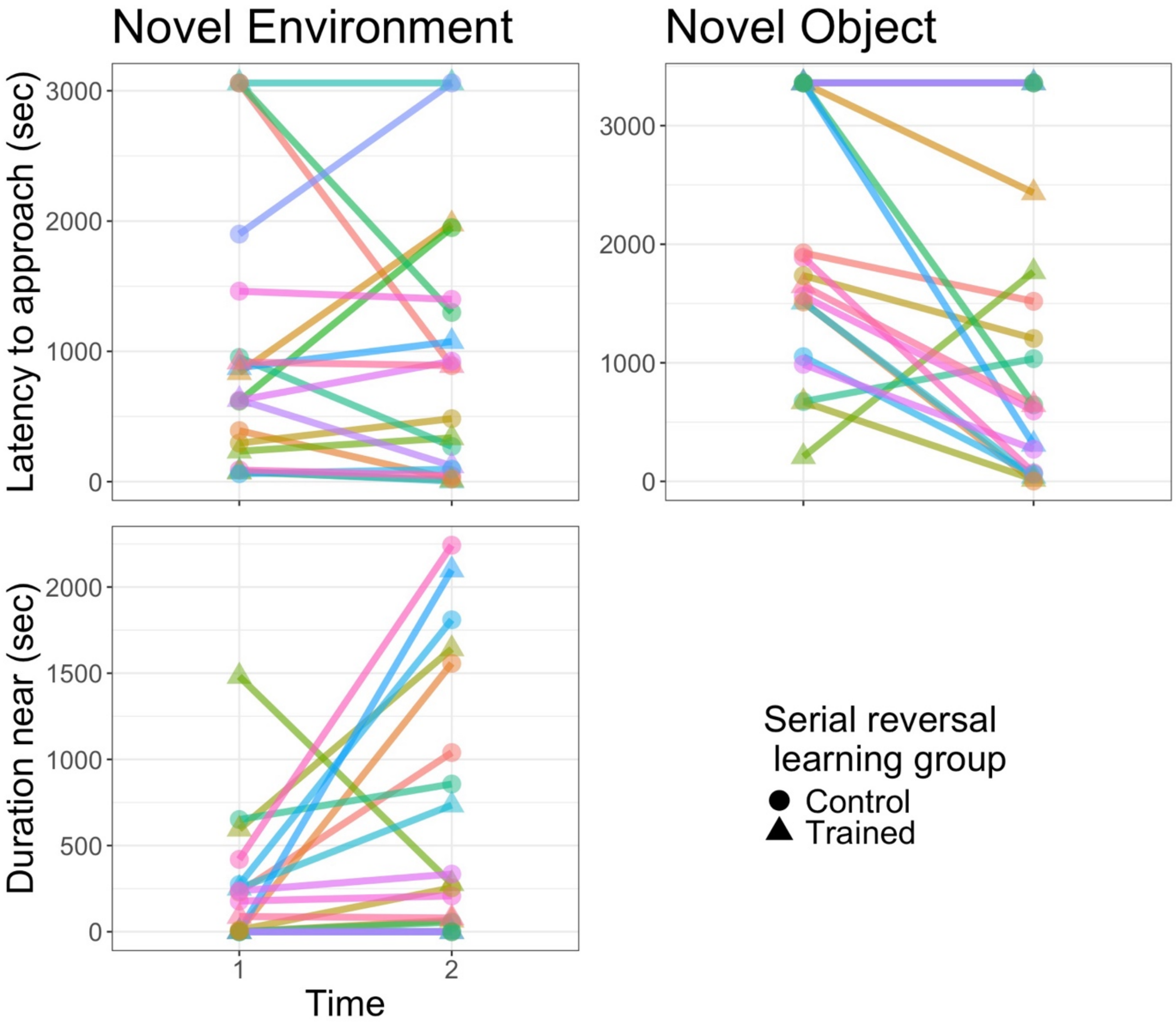

Figure 3 - The latency to approach and the duration spent near the novel environment test were significantly repeatable across time, whereas performance was not repeatable for novel object exploration. Each line color represents an individual and the points show the amount of time before individuals approached to within 100 cm (Latency to approach) or amount of time individuals spent within 20cm (Duration near) of the novel item during each of the two 45-minute trials. The shape of the point is based on whether the grackle was part of the control (circle) or trained (triangle) group in the serial reversal learning experiment. The two time periods were separated by 34 days on average (range: 11-49). If performance is repeatable within a test we would expect the line connecting the two points to be at or close to horizontal, and the lines of different individuals to be approximately parallel.

Persistence

We tested whether individuals (n = 17) were repeatable in the number of touches per trial time that they made across multiple novel test apparatuses (Figure 1b-f): boldness objects, exploration environment and object, as well as the two different MABs. We found that total number of touches per trial time across these diverse objects was repeatable (R = 0.28, CI = 0.07-0.46, $p < 0.01$; Figure 4). However, this was driven by functional touches because touches to the MABs that were nonfunctional (i.e., applied to the parts of the apparatus that could never result in obtaining the food) were not repeatable (R = 0.08, CI = 0.00 - 0.58, $p = 1$).

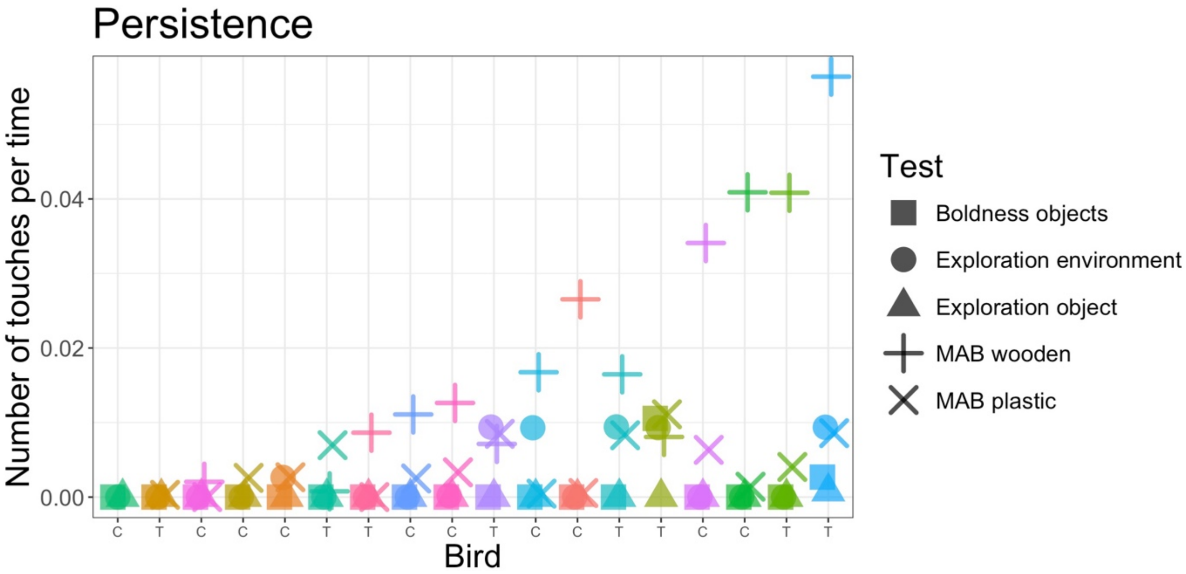

Figure 4 - Persistence (the number of touches per time) was repeatable across multiple diverse test apparatuses. The x-axis shows each individual bird, also identified by unique colors and labeled with a “C” or a “T” to distinguish whether they were part of the control or trained group, respectively, in the serial reversal learning experiment. Birds are sorted on the x-axis according to the maximum number of touches per time. Test apparatuses are distinguished by shape and we abbreviated multiaccess box as “MAB” in the figure legend.

Motor Diversity

We quantified the number of different motor behaviors used while interacting with two distinct MABs in 17 grackles. Grackles were not consistent in the number of motor behaviors used across the two MABs and so repeatability was very low and not statistically significant (R = 0.06, CI = 0.00-0.45, $p = 0.50$).

Relationships among measures

The repeatability analyses informed which of our methods measured consistent individual differences in behavior. Our next goal was to investigate the relationships among only the repeatable measures (exploration of a novel environment and persistence) and the Flexibility Comprehensive variables and whether the grackle was in the flexibility trained or control group.

Relationship between flexibility and exploration

We first analyzed the relationship between the Flexibility Comprehensive measures that quantify the rate of learning to be attracted to a color option in the serial reversal learning task, ϕ, and the rate of deviating from learned associations, λ (Blaisdell et al., 2021; Lukas et al., 2024), and two variables describing novel environment exploration: Duration near (within 20cm) the outside of the tent, and the latency to first come to the ground from the aviary perches to approach the tent. We found no relationship between either measure of novel environment exploration and ϕ or λ (Table 2).

We next investigated if performance varied as a function of whether individuals went through serial reversal learning to increase flexibility (trained group, n=8) or not (control group, n=11). Grackles that underwent the flexibility training were more exploratory in that they spent more time within 20cm of the outside of the novel environment relative to control individuals (Table 3; *ß* = 3.92, $p = 0.04$). However, there was no difference between trained and control individuals in latency to come to the ground within 100 cm of the novel environment (*ß* = -0.43, $p = 0.54$).

Table 2 - Behavioral flexibility, measured with two variables comprising our Flexibility Comprehensive measure (ϕ - the learning rate of attraction to either option, and λ - the rate of deviating from learned attractions), was not related to exploratory tendency as measured by duration spent within 20cm of the outside of the novel environment (Duration near) or the latency to approach to within 100 cm of the novel object (Latency to land). Moreover, persistence (Number of functional touches per time) was not related to ϕ or λ.

	Duration near			Latency to land			Number of functional touches		
	Est. (CI)	S.E.	*p*	Est. (CI)	S.E.	*p*	Est. (CI)	S.E.	*p*
Intercept	2.68 (1.09, 4.26)	0.81	<0.01	6.07 (5.38, 6.77)	0.34	<0.01	-6.10 (-6.51, -5.68)	0.21	<0.01
c.phi[1]	1.50 (-0.18, 3.17)	0.86	0.08	-0.56 (-1.31, 0.19)	0.37	0.14	0.27 (-0.19, 0.73)	0.23	0.25
c.lambda[1]	-0.22 (-1.86, 1.42)	0.84	0.79	0.02 (-0.73, 0.77)	0.37	0.96	-0.09 (-0.54, 0.36	0.23	0.71

[1] ‘c.phi’ and ‘c.lambda’ represent the centered and scaled version of these variables because the phi and lambda values were on fairly different scales.

Table 3 - We assessed whether exploration and persistence were related to behavioral flexibility training. Grackles in the trained group that were more behaviorally flexible in the serial reversal learning task were more exploratory in that they spent more time within 20cm of the outside of the tent compared to control individuals. Whereas, latency to land within 100 cm of the novel environment and persistence were not related to behavioral flexibility training.

	Duration near			Latency to land			Number of functional touches		
	Est. (CI)	S.E.	*p*	Est. (CI)	S.E.	*p*	Est. (CI)	S.E.	*p*
Intercept	1.16 (-0.87, 3.19)	1.04	0.26	6.26 (5.31, 7.21)	0.47	<0.01	-6.37 (-6.93, -5.81)	0.29	<0.01
Flexibility trained	3.61 (0.56, 6.67)	1.56	**0.02**	-0.44 (-1.90, 1.03)	0.72	0.55	0.62 (-0.20, 1.44)	0.42	0.14

Relationship between flexibility and persistence

We found no support for a relationship between persistence, measured as functional touches to all test apparatuses, and either ϕ (Table 2; n=19, *ß* = 0.42, $p = 0.11$) or λ (*ß* = 0.08, $p = 0.77$). We then looked at whether the number of incorrect choices in the reversal learning task (i.e., how much the grackle is perseverating on a previously rewarded color option before exploring the other option, which is considered a measure of persistence) was related to the average number of functional or nonfunctional touches per time to the novel apparatuses (see P3 alternative 2, above). We found no evidence of a relationship between these two potential measures of persistence because the intercept-only model was supported over the model containing the number of touches variable (Table S2). This is evidence that the number of touches is not related to perseverating on an option in a way that inhibits flexible learning.

Lastly, in contrast to the exploration results, we found no evidence of a relationship between persistence and whether or not the grackle underwent the flexibility training. The number of functional touches to the novel apparatuses did not differ between control and trained grackles (Table 3; *ß* = 0.81, $p = 0.09$).

Discussion

Rapid human-induced environmental change leads to novel challenges for wildlife, where individual and species ability to survive is most often possible through behavioral change (Wright et al., 2010). Although several behavioral traits are implicated in successful adaptation to human modified environments (Chapple et al., 2012), it is uncommon to directly test for multiple traits in the same individuals. Here, we used multiple novel and threatening stimuli to assess the validity of methods measuring various behavioral traits, and the relationships among traits, in great-tailed grackles, a species that has adapted to many human-induced changes to its environment during a rapid range expansion. We found that only some of our methods for measuring behavioral traits in captivity produced repeatable performance and in support of our main hypothesis, we did find a relationship between behavioral flexibility and exploration.

Personality traits like boldness, exploration, and persistence are not directly observable. To validate that the experimental method used likely elicited performance reflective of the inherent personality trait, performance must be repeatable across time and contexts (Carter et al., 2013). We found that the number of total touches (functional and nonfunctional) that grackles made to multiple different novel apparatuses was repeatable, indicating that this is likely a valid method for measuring the trait persistence. Despite using multiple assays and stimuli to quantify exploration, boldness, and motor diversity, we found that only one method produced repeatable performance: the novel environment exploration assay. The other methods, exploration of a novel object, boldness towards two different novel threats, and the number of distinct motor behaviors used to interact with the two different MABs (Figure 1) did not produce repeatable performance across sampling periods. However, we provide in Supplementary Material 2 a plot of the raw boldness and exploration data so readers can visually compare performance among tests (Figure S2).

A key aspect distinguishing boldness from exploration is that boldness reflects a response to potentially threatening objects, novel or familiar (Carter et al., 2013; Greggor et al., 2015). Consequently, we compared performance between the novel or threatening objects and the familiar objects in the exploration and boldness assays. The novel environment was the only object the grackles did not perceive as a threat. Although the novel object for the exploration assay was not meant to be threatening (e.g., it was smaller than the threatening objects, it did not have eyes), grackles still spent significantly less time near it than their familiar object. Consequently, grackles did not perform consistently on these assays where the object was perceived as threatening. This highlights the relevance of the jingle-jangle fallacy, which describes the mismatch between a trait label and what the method actually measures (Carter et al., 2013). Although we expected the novel object to measure the trait exploration, by incorporating control conditions and multiple other novel and threatening objects, it was clear that the novel object was eliciting performance that was likely more reflective of boldness.

It is possible that grackles, in general, do not produce repeatable responses when faced with a threat in captivity. In the wild, grackles are a gregarious species that probably rarely encounters threats while alone (Johnson & Peer, 2022). For several reasons, we did not house more than one grackle in each aviary. Therefore, the lack of repeatability in performance could stem from the relatively contrived situation of experiencing a threat when visually isolated from conspecifics. This preliminary evidence is congruent with other research on social species encountering novelty. For example, zebra finches were more likely to approach a novel object for food (Coleman & Mellgren, 1994) and investigate a novel environment (Schuett & Dall, 2009) when in a social group compared to when alone. However, Carib grackles were slower to approach novel foraging opportunities when in a social group compared to when alone (Overington et al., 2009). Because the majority of research on animal personality traits is conducted on individuals in captivity regardless of their sociality, more research is needed to understand when social behavior may affect the consistency of performance on personality assays.

We assessed the relationship between our repeatable behavioral traits (exploration and persistence) and the two measures of behavioral flexibility (Flexibility Comprehensive and flexibility trained versus control groups). Our Flexibility Comprehensive measure reflects two aspects of performance during serial reversal learning, the rate of learning to be attracted to a color option, ϕ, and the rate of deviating from learned associations, λ (Blaisdell et al., 2021; Lukas et al., 2024). We predicted that exploration would be positively related to flexibility, and in particular we assumed λ would best reflect exploratory behavior during the reversal learning task (Lukas et al., 2024). We found no relationship between the Flexibility Comprehensive variables and novel environment exploration. This is contrary to previous literature that found that flexibility is theoretically (Griffin et al., 2016) and experimentally (Rojas-Ferrer et al., 2020) linked with this behavioral trait. However, in support of previous literature, we found that grackles that underwent the serial reversal learning training to experimentally increase flexibility were more exploratory towards the novel environment compared to grackles that were in the control group. This potentially explains how great-tailed grackles are successful at adapting to rapid anthropogenic change. The individuals in the population that are willing to seek out novel foraging or nesting opportunities are also able to change their behavior to switch to using these novel resources when they are encountered.

The inconsistent results for the relationship of exploration with either of the two different measures of flexibility likely reflects that individuals trained to be more flexible through serial reversal learning ended up with different strategies for how to reverse quickly (Lukas et al., 2024). Trained individuals had a higher ϕ and lower λ relative to grackles in the control group. As such, trained individuals were good at reacting to changes in the environment either because they kept on exploring alternative options (high λ) or because they placed high importance on new information (high ϕ). With either strategy, we could expect trained individuals to also be better at exploration. In addition, we found that, even though all grackles improved during the training, individual differences persisted (McCune et al., 2023). These individual differences might be linked to their persistence, which would explain why the training did not influence the relationship between flexibility and persistence.

In addition, with a sample size of 19, we potentially lacked the power to detect a subtle relationship between flexibility and exploration or persistence. We conducted a power analysis a priori that indicated that a sample size of 32 would permit detections of large effect sizes. We did not meet this sample size goal, due to the difficulty in catching grackles and the large time commitment for serial reversal learning, and so it is possible we failed to detect some relationships. However, the power analysis included many more predictor variables than we ended up using (see Changes after the study began) and was conducted before we determined that the serial reversal learning trained grackles to be significantly more flexible than control grackles (Logan et al., 2023). Thus, the increased difference in flexibility between control and trained grackles, also reflected in the ϕ and λ values, should increase our power to detect a relationship between these behavioral traits and flexibility, if it exists. Nevertheless, future research should evaluate these relationships with larger sample sizes.

By assessing multiple behavioral traits in the same individuals of a highly adaptable species, we were able to identify correlations among certain repeatable traits that can inform our understanding of the ability to adapt to environmental change. Overall, we found that the time spent exploring near a novel environment is related to flexibility. Our results support previous hypotheses about traits that are related to flexible behavior, and therefore might be important for increasing survival and fitness in the face of human-induced environmental change. However, additional research is needed to further validate methods for measuring individual differences in boldness and motor diversity in this species, and to disentangle the mechanisms driving the mixed results for the relationship between persistence, exploration, and the two ways of measuring behavioral flexibility.

Acknowledgements

Preprint version 4 of this article has been peer-reviewed and recommended by Peer Community In Ecology (https://doi.org/10.24072/pci.ecology.100771; Van Cleve, 2025). We thank Jeremy Van Cleve and two anonymous reviewers for their feedback on the Stage 1 preregistration and the additional two anonymous reviewers for feedback at Stage 2. We are grateful to Julia

Cissewski for tirelessly solving problems involving financial transactions and contracts; Sophie Kaube for logistical support; and Richard McElreath for project support. Addditionally, thanks to Kevin Langergraber for serving as local PI on the Arizona State University IACUC; Ben Trumble for providing us with a wet lab at ASU; Melissa Wilson Sayres for sponsoring our affiliations at ASU and lending lab equipment; Kristine Johnson for technical advice on great-tailed grackles; ASU School of Life Sciences Department Animal Care and Technologies for providing space for our aviaries and for their excellent support of our daily activities; and our research assistants who helped habituate, trap, weigh, and train grackles to participate in tests: Nancy Rodriguez, Aelin Mayer, Sofija Savic, Brianna Thomas, Aldora Messinger, Elysia Mamola, Michael Guillen, Rita Barakat, Adrianna Boderash, Olateju Ojekunle, August Sevchick, and Justin Huynh. We thank Melissa Folsom and Luisa Bergeron for their exceptional assistance with logistical planning and data collection in the field.

Author Contributions

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

Ethics

The research on the great-tailed grackles followed established ethical guidelines for the involvement and treatment of animals in experiments and received institutional approval prior to conducting the study (US Fish and Wildlife Service scientific collecting permit number MB76700A-0,1,2; US Geological Survey Bird Banding Laboratory federal bird banding permit number 23872; Arizona Game and Fish Department scientific collecting license number SP594338 [2017], SP606267 [2018], and SP639866 [2019]; Institutional Animal Care and Use Committee at Arizona State University protocol number 17-1594R; University of Cambridge ethical review process non-regulated use of animals in scientific procedures: zoo4/17 [2017]).

Funding

This research is funded by the Department of Human Behavior, Ecology and Culture at the Max Planck Institute for Evolutionary Anthropology, and by a Leverhulme Early Career Research Fellowship to Logan in 2017-2018.

Conflict of interest disclosure

The authors declare that they comply with the PCI rule of having no financial conflicts of interest in relation to the content of the article. The authors declare the following non-financial conflict of interest: Logan and Lukas are Recommenders at PCI Ecology and PCI Registered Reports, and Logan was on the Managing Board at PCI Ecology (2018-2022) and is on the Managing Board at PCI Registered Reports (2021-current).

Data, scripts, code, and supplementary information availability

All data and code to reproduce the analysis included in this manuscript can be found at the KNB data repository (McCune & Logan, 2025): https://doi.org/10.5063/F118350Q. Supplementary information is available on OSF (McCune, 2025): https://doi.org/10.17605/OSF.IO/2UNGZ

References

Amy M, van Oers K, Naguib M (2012) Worms under cover: relationships between performance in learning tasks and personality in great tits (Parus major). Animal Cognition, 15, 763–770. https://doi.org/10.1007/s10071-012-0500-3

Bebus SE, Small TW, Jones BC, Elderbrock EK, Schoech SJ (2016) Associative learning is inversely related to reversal learning and varies with nestling corticosterone exposure. Animal Behaviour, 111, 251–260. https://doi.org/10.1016/j.anbehav.2015.10.027

Bensky MK, Bell AM (2022) A Behavioral Syndrome Linking Boldness and Flexibility Facilitates Invasion Success in Sticklebacks. The American Naturalist, 200, 846–856. https://doi.org/10.1086/721765

Biro PA, Dingemanse NJ (2008) Sampling bias resulting from animal personality. Trends in Ecology and Evolution, 24, 66–67. https://doi.org/10.1016/j.tree.2008.11.001

Blaisdell A, Seitz B, Rowney C, Folsom M, MacPherson M, Deffner D, Logan CJ (2021) Do the more flexible individuals rely more on causal cognition? Observation versus intervention in causal inference in great-tailed grackles. Peer Community Journal, 1, e50. https://doi.org/10.24072/pcjournal.44

Brehm AM, Mortelliti A (2018) Mind the trap: Large-scale field experiment shows that trappability is not a proxy for personality. Animal Behaviour, 142, 101–112. https://doi.org/10.1016/j.anbehav.2018.06.009

Canestrelli D, Bisconti R, Carere C (2016) Bolder takes all? The behavioral dimension of biogeography. Trends in Ecology & Evolution, 31, 35–43. https://doi.org/10.1016/j.tree.2015.11.004

Carter AJ, Feeney WE, Marshall HH, Cowlishaw G, Heinsohn R (2013) Animal personality: What are behavioural ecologists measuring? Biological Reviews, 88, 465–475. https://doi.org/10.1111/brv.12007

Chapple DG, Simmonds SM, Wong BBM (2012) Can behavioral and personality traits influence the success of unintentional species introductions? Trends in Ecology and Evolution, 27, 57–62. https://doi.org/10.1016/j.tree.2011.09.010

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

van Cleve J (2025) Exploring exploration and behavioral flexibility in grackles: how to handle issues of “jingle-jangle” and repeatability. Peer Community in Ecology, 100771. https://doi.org/10.24072/pci.ecology.100771

Coleman SL, Mellgren RL (1994) Neophobia when feeding alone or in flocks in zebra finches, Taeniopygia guttata. Animal Behaviour, 48, 903–907. https://doi.org/10.1006/anbe.1994.1315

Croston R, Branch CL, Pitera AM, Kozlovsky DY, Bridge ES, Parchman TL, Pravosudov VV (2017) Predictably harsh environment is associated with reduced cognitive flexibility in wild food-caching mountain chickadees. Animal Behaviour, 123, 139–149. https://doi.org/10.1016/j.anbehav.2016.10.004

De Meester G, Pafilis P, Van Damme R (2022) Bold and bright: shy and supple? The effect of habitat type on personality–cognition covariance in the Aegean wall lizard (Podarcis erhardii). Animal Cognition, 25, 745–767. https://doi.org/10.1007/s10071-021-01587-0

Dingemanse NJ, Dochtermann NA (2013) Quantifying individual variation in behaviour: Mixed-effect modelling approaches. Journal of Animal Ecology, 82, 39–54. https://doi.org/10.1111/1365-2656.12013

Diquelou MC, Griffin AS, Sol D (2016) The role of motor diversity in foraging innovations: A cross-species comparison in urban birds. Behavioral Ecology, 27, 584–591. https://doi.org/10.1093/beheco/arv190

Dougherty LR, Guillette LM (2018) Linking personality and cognition: a meta-analysis. Philosophical Transactions of the Royal Society B: Biological Sciences, 373, 20170282. https://doi.org/10.1098/rstb.2017.0282

Duckworth RA (2010) Evolution of Personality: Developmental Constraints on Behavioral Flexibility. The Auk, 127, 752–758. https://doi.org/10.1525/auk.2010.127.4.752

Fidler AE, Van Oers K, Drent PJ, Kuhn S, Mueller JC, Kempenaers B (2007) Drd4 gene polymorphisms are associated with personality variation in a passerine bird. Proceedings of the Royal Society B: Biological Sciences, 274, 1685–1691. https://doi.org/10.1098/rspb.2007.0337

Greggor AL, Thornton A, Clayton NS (2015) Neophobia is not only avoidance: Improving neophobia tests by combining cognition and ecology. Current Opinion in Behavioral Sciences, 6, 82–89. https://doi.org/10.1016/j.cobeha.2015.10.007

Griffin AS, Guez D, Federspiel I, Diquelou MC, Lermite F (2016) Ch. 3 Invading new environments: a mechanistic framework linking motor diversity and cognition to establishment success. In: Biological Invasions and Animal Behaviour , pp. 26–46. Cambridge University Press. https://doi.org/10.1017/cbo9781139939492.004.

Griffin AS, Guillette LM, Healy SD (2015) Cognition and personality: An analysis of an emerging field. Trends in Ecology & Evolution, 30, 207–214. https://doi.org/10.1016/j.tree.2015.01.012

Guenther A, Brust V, Dersen M, Trillmich F (2014) Learning and personality types are related in cavies (Cavia aperea). Journal of Comparative Psychology, 128, 74–81. https://doi.org/10.1037/a0033678

Hadfield JD (2010) MCMCglmm: MCMC Methods for Multi-Response GLMMs in R. Journal of Statistical Software, 33, 1–22. https://doi.org/10.18637/jss.v033.i02

Hartig F (2022) DHARMa: Residual diagnostics for hierarchical (multi-level/mixed) regression models. https://doi.org/10.32614/CRAN.package.DHARMa

Hendry AP, Farrugia TJ, Kinnison MT (2008) Human influences on rates of phenotypic change in wild animal populations. Molecular Ecology, 17, 20–29. https://doi.org/10.1111/j.1365-294X.2007.03428.x

Johnson K, Peer BD (2022) Great-tailed Grackle (Quiscalus mexicanus). In: Birds of the World (P.G. Rodewald and B.K. Kenney, Editors) , pp. 1–28. Cornell Lab of Ornithology, Ithaca, NY, USA. https://doi.org/10.2173/bow.grtgra.02

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

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

Logan C, Lukas D, Blaisdell A, Seitz B, Johnson-Ulrich Z, MacPherson M, Sevchik A, McCune KB (2023) Behavioral flexibility is manipulable and it improves flexibility and innovativeness in a new context. Peer Community Journal, 3. https://doi.org/10.24072/pcjournal.284

Logan C, Lukas D, Geng X, LeGrande-Rolls C, Marfori Z, MacPherson M, Rowney C, Smith C, McCune K (2025) Behavioral flexibility is related to foraging, but not social or habitat use behaviors, in a species that is rapidly expanding its range. Peer Community Journal, 5, e74. https://doi.org/10.32942/X2T036

Lukas D, McCune K, Blaisdell A, Johnson-Ulrich Z, MacPherson M, Seitz B, Sevchik A, Logan C (2024) Bayesian reinforcement learning models reveal how great-tailed grackles improve their behavioral flexibility in serial reversal learning experiments. Peer Community Journal, 4, e88. https://doi.org/10.24072/pcjournal.456

McCune KB (2025) Supplementary material for “Behavioral flexibility is related to exploration, but not boldness, persistence or motor diversity.” https://doi.org/10.17605/OSF.IO/2UNGZ

McCune KB, Blaisdell AP, Johnson-Ulrich Z, Sevchik A, Lukas D, MacPherson M, Seitz BM, Logan C (2023) Using repeatability of performance within and across contexts to validate measures of behavioral flexibility. PeerJ, 11, 1–17. https://doi.org/10.7717/peerj.15773

McCune KB, Jablonski PG, Lee S, Ha RR (2018) Evidence for personality conformity, not social niche specialization in social jays. Behavioral Ecology, 29, 910–917. https://doi.org/10.1093/beheco/ary055

McCune KB, Logan CJ (2025) Data: Behavioral flexibility is related to exploration, but not boldness, persistence or motor diversity. https://doi.org/doi:10.5063/F118350Q

Mettke-Hofmann C, Winkler H, Leisler B (2002) The significance of ecological factors for exploration and neophobia in parrots. Ethology, 108, 249–272. https://doi.org/10.1046/j.1439-0310.2002.00773.x

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

Morand-Ferron J, Reichert MS, Quinn JL (2022) Cognitive flexibility in the wild: Individual differences in reversal learning are explained primarily by proactive interference, not by sampling strategies, in two passerine bird species. Learning and Behavior, 50, 153–166. https://doi.org/10.3758/s13420-021-00505-1

Nakagawa S, Schielzeth H (2010) Repeatability for Gaussian and non-Gaussian data: A practical guide for biologists. Biological Reviews, 85, 935–956. https://doi.org/10.1111/j.1469-185X.2010.00141.x

Overington SE, Cauchard L, Morand-Ferron J, Lefebvre L (2009) Innovation in groups: does the proximity of others facilitate or inhibit performance? Behaviour, 146, 1543–1564. https://doi.org/10.1163/156853909X450131

Perals D, Griffin AS, Bartomeus I, Sol D (2017) Revisiting the open-field test: what does it really tell us about animal personality? Animal Behaviour, 123, 69–79. https://doi.org/10.1016/j.anbehav.2016.10.006

Plummer M, Best N, Cowles K, Vines K, Sarkar D, Bates D, Almond R, Magnusson A (2024) coda: Output analysis and diagnostics for MCMC. R News, 6, 7–11. https://doi.org/10.32614/CRAN.package.coda

Reale D, Reader SM, Sol D, McDougall PT, Dingemanse NJ (2007) Integrating animal temperament within ecology and evolution. Biological Reviews, 82, 291–318. https://doi.org/10.1111/j.1469-185X.2007.00010.x

Rojas-Ferrer I, Thompson MJ, Morand-Ferron J (2020) Is exploration a metric for information gathering? Attraction to novelty and plasticity in black-capped chickadees. Ethology, 126, 383–392. https://doi.org/10.1111/eth.12982

Rowe C, Healy SD (2014) Measuring variation in cognition. Behavioral Ecology, 25, 1287–1292. https://doi.org/10.1093/beheco/aru090

Schuett W, Dall SRX (2009) Sex differences, social context and personality in zebra finches Taeniopygia guttata. Animal Behaviour, 77, 1041–1050. https://doi.org/10.1016/j.anbehav.2008.12.024

Shaw RC, Schmelz M (2017) Cognitive test batteries in animal cognition research: evaluating the past, present and future of comparative psychometrics. Animal Cognition, 20, 1003–1018. https://doi.org/10.1007/s10071-017-1135-1

Sih A (2013) Understanding variation in behavioural responses to human-induced rapid environmental change: A conceptual overview. Animal Behaviour, 85, 1077–1088. https://doi.org/10.1016/j.anbehav.2013.02.017

Sih A, Bell AM, Johnson JC (2004) Behavioral syndromes: An ecological and evolutionary overview. Trends in Ecology and Evolution, 19, 372–378. https://doi.org/10.1016/j.tree.2004.04.009

Sih A, Ferrari MCO, Harris DJ (2011) Evolution and behavioural responses to human-induced rapid environmental change. Evolutionary Applications, 4, 367–387. https://doi.org/10.1111/j.1752-4571.2010.00166.x

Sol D, Lapiedra O, González-Lagos C (2013) Behavioural adjustments for a life in the city. Animal Behaviour, 85, 1101–1112. https://doi.org/10.1016/j.anbehav.2013.01.023

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

Stoffel MA, Nakagawa S, Schielzeth H (2017) rptR: Repeatability estimation and variance decomposition by generalized linear mixed-effects models. Methods in Ecology and Evolution, 8, 1639–1644. https://doi.org/10.1111/2041-210X.12797

Summers J, Lukas D, Logan CJ, Chen N (2023) The role of climate change and niche shifts in divergent range dynamics of a sister-species pair. Peer Community Journal, 3, e23. https://doi.org/10.24072/pcjournal.248

Takola E, Krause ET, Müller C, Schielzeth H (2021) Novelty at second glance: a critical appraisal of the novel object paradigm based on meta-analysis. Animal Behaviour, 180, 123–142. https://doi.org/10.1016/j.anbehav.2021.07.018

Titulaer M, van Oers K, Naguib M (2012) Personality affects learning performance in difficult tasks in a sex-dependent way. Animal Behaviour, 83, 723–730. https://doi.org/10.1016/j.anbehav.2011.12.020

Wehtje W (2003) The range expansion of the great-tailed grackle (Quiscalus mexicanus Gmelin) in North America since 1880. Journal of Biogeography, 30, 1593–1607. https://doi.org/10.1046/j.1365-2699.2003.00970.x

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

Zhang X, Yi N (2020) NBZIMM: Negative binomial and zero-inflated mixed models, with application to microbiome/metagenomics data analysis. BMC Bioinformatics, 21, 1–19. https://doi.org/10.1186/s12859-020-03803-z