

Is behavioral flexibility linked with exploration, but not boldness, persistence, or motor diversity?

Dr. Kelsey McCune (University of California Santa Barbara / Max Planck Institute for Evolutionary Anthropology), Carolyn Rowney (Max Planck Institute for Evolutionary Anthropology), Luisa Bergeron (University of California Santa Barbara / Max Planck Institute for Evolutionary Anthropology), Dr. Corina Logan (<http://CorinaLogan.com>) (Max Planck Institute for Evolutionary Anthropology, corina_logan@eva.mpg.de)

27 March 2019

```
library(knitr)
opts_chunk$set(tidy.opts=list(width.cutoff=60),tidy=TRUE)
```

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

Jeremy Van Cleve (2019) Probing behaviors correlated with behavioral flexibility. *Peer Community in Ecology*, 100020.

10.24072/pci.ecology.100020 (<https://ecology.peercommunityin.org/public/rec?id=29&reviews=True>) - Reviewers: two anonymous reviewers

Cite as: McCune K, Rowney C, Bergeron L, Logan CJ. 2019. Is behavioral flexibility linked with exploration, but not boldness, persistence, or motor diversity? (http://corinalogan.com/Preregistrations/g_exploration.html) (http://corinalogan.com/Preregistrations/g_exploration.html (http://corinalogan.com/Preregistrations/g_exploration.html)) In principle acceptance by *PCI Ecology* of the version on 27 Mar 2019
https://github.com/corinalogan/grackles/blob/a4ef45b5225fd2ae937202333d9e0c5c75e76b88/Files/Preregistrations/g_exploration.Rmd (https://github.com/corinalogan/grackles/blob/a4ef45b5225fd2ae937202333d9e0c5c75e76b88/Files/Preregistrations/g_exploration.Rmd).

ABSTRACT

This is one of the first studies planned for our long-term research on the role of behavioral flexibility in rapid geographic range expansions. **Project background:** Behavioral flexibility, the ability to change behavior when circumstances change based on learning from previous experience (Mikhalevich, Powell, and Logan (2017)), is thought to play an important role in a species' ability to successfully adapt to new environments and expand its geographic range (e.g., (Lefebvre et al. 1997), (Griffin and Guez 2014), (Chow, Lea, and Leaver 2016), (Sol and Lefebvre 2000), (Sol, Timmermans, and Lefebvre 2002), (Sol et al. 2005)). However, behavioral flexibility is rarely directly tested at the individual level, thus limiting our ability to determine how it relates to other traits, which limits the power of predictions about a species' ability to adapt behavior to new environments. We use great-tailed grackles (a bird species) as a model to investigate this question because they have rapidly expanded their range into North America over the past 140 years ((Wehtje 2003), (Peer 2011)) (see an overview of the 5-year project timeline ([./README.md](#))). **This investigation:** In this piece of the long-term project, we aim to understand whether grackle behavioral flexibility (color tube reversal learning - described in a separate preregistration (https://github.com/corinalogan/grackles/blob/master/EasyToReadFiles/g_flexmanip.md)) correlates (or not) with individual differences in the exploration of new environments and novel objects, boldness, persistence, and motor diversity (and whether the flexibility manipulation made such correlations more detectable). Results will indicate whether consistent individual differences in these traits might interact with measures of flexibility (reversal learning and solution switching). This will improve our understanding of which variables are linked with flexibility and how they are related, thus putting us in an excellent position to further investigate the mechanisms behind these links in future research.

A. STATE OF THE DATA

Prior to collecting any data: This preregistration was written and submitted to PCI Ecology for peer review (Sep 2018).

After data collection had begun (and before any data analysis was conducted): This preregistration was peer reviewed at PCI Ecology, revised, and resubmitted (Feb 2019), and passed pre-study peer review (Mar 2019). See the peer review history (<https://ecology.peercommunityin.org/public/rec?id=29&reviews=True>).

B. PARTITIONING THE RESULTS

We may decide to present the results from different hypotheses in separate papers.

C. HYPOTHESES

H1: Behavioral flexibility

(https://github.com/corinalogan/grackles/blob/master/EasyToReadFiles/g_flexmanip.md) (indicated by individuals that are faster at functionally changing their behavior when circumstances change; measured by reversal learning and switching between options on a multi-access box) is positively correlated with the exploration of new environments and novel objects, but not with other behaviors (i.e., boldness, persistence, or motor diversity) (see Mikhalevich, Powell, and Logan (2017) for theoretical background about our flexibility definition).

We will first verify that our measures of exploration, boldness and persistence represent repeatable, inherent individual differences in behavior (i.e., personality). Individuals show consistent individual differences in behavior if the variance in latency to approach the task is smaller within individuals compared to variance in latency among individuals (for exploration and boldness assays). The same definition applies to persistence with the number of touches as the measured variable. If there is no repeatability of these behaviors within individuals, then performance is likely state dependent (e.g., it depends on their fluctuating motivation, hunger levels, etc.) and/or reliant on the current context of the tasks.

Predictions 1-5: Individuals in the experimental group where flexibility (as measured by reversal learning and on a multi-access box) was manipulated (such that individuals in the manipulated group became faster at switching) will be more exploratory of new environments (P1; methods similar to free-entry open field test as in Mettke-Hofmann et al. (2009)) and novel objects (P2; methods as in Mettke-Hofmann et al. (2009)) than individuals in the control group where flexibility was not increased, and there will be no difference between the groups in persistence (P3), boldness (P4; methods as in Logan (2016)), or motor diversity (P5) (as found in Logan (2016)). We do not expect the flexibility manipulation to causally change the nature of the relationship between flexibility and any of the other measured variables. Instead, we expect the manipulation to potentially enhance individual variation, thus making it easier for us to detect a correlation if one exists.

P1-P5 alternative: If the flexibility manipulation does not work in that those individuals in the experimental condition are not more flexible than control individuals, then we will analyze the individuals from both conditions as one group. In this case, we will assume that we were not able to influence their flexibility and that whatever level of flexibility they had coming into the experiment reflects the general individual variation in the population. This experiment will then elucidate whether general individual variation in flexibility relates to exploratory behaviors. The predictions are the same as above. The following alternatives apply to both cases: if the manipulation works (in which case we expect stronger effects for the manipulated group), and if the manipulation doesn't work (in which case we expect individuals to vary across all of the measured variables and for these variables to potentially interact).

P1 alternative 1: There is a positive correlation between exploration and both dependent variables in reversal learning (one accounts for exploration in reversal learning [the ratio] and the other does not). This suggests that flexibility is not independent of exploration and could indicate that another trait is present that could be explaining individual variation in flexibility as well as in exploration. This other trait or traits could be something such as boldness or persistence.

P1 alternative 2a: There is a positive correlation between exploration and the dependent variable that does not account for exploration (number of trials to reverse), but not the flexibility ratio, which suggests that performance overall in reversal learning is partially explained by variation in exploration, but that flexibility and exploration are separate traits because using a measure that accounts for exploration still shows variation in flexibility.

P1 alternative 2b: There is a negative correlation between exploration and the flexibility ratio that accounts for exploration, but not with the number of trials to reverse. This could be an artifact of accounting for exploration in both variables.

P1 alternative 3: There is no correlation between exploration and either dependent variable in reversal learning. This indicates that both dependent variables measure traits that are independent of exploration.

P1 alternative 4: There is no correlation between exploration and either dependent variable in reversal learning 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 is 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.

Does manipulating flexibility affect...

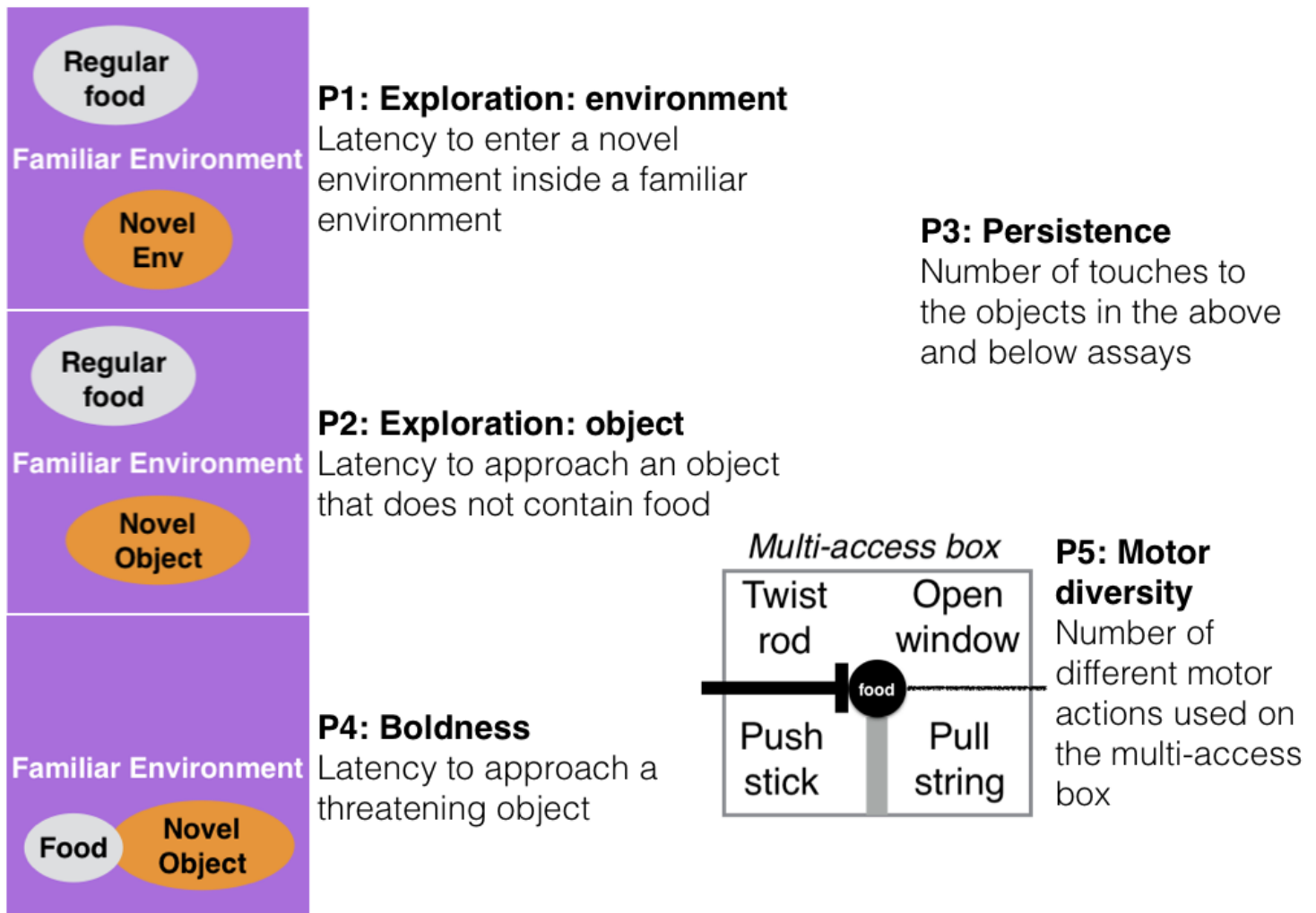


Figure 1.

Figure 1. An overview of the study design and a selection of the variables we will measure for each assay. Exploration will be measured by comparing individual behavior within a familiar environment to behavior towards a novel environment, as well as response to a familiar object vs. a novel object within the familiar environment that contains their regular food. Boldness will be measured as the willingness to eat next to a threatening object (familiar, novel object, or a taxidermic predator) in their familiar environment. Persistence will be measured as the number of touches to the novel environment and novel object in the Exploration assay, the objects in the Boldness assay, and the multi-access box in a separate preregistration

(https://github.com/corinalogan/grackles/blob/master/EasyToReadFiles/g_flexmanip.md). Motor diversity will be measured using the multi-access box in a separate preregistration

(https://github.com/corinalogan/grackles/blob/master/EasyToReadFiles/g_flexmanip.md). After the flexibility manipulation occurs, assays will be conducted at least twice (e.g., Time 1, Time 2) and differences (if any) between the control and manipulated groups in the behavioral flexibility preregistration (https://github.com/corinalogan/grackles/blob/master/EasyToReadFiles/g_flexmanip.md) will be compared across time and, with persistence, across tests (e.g., Test 1, Test 2) because persistence is measured in four different assays.

H2: Captive and wild individuals may respond differently to assays measuring exploration and boldness.

P6: Individuals assayed while in captivity are less exploratory and bold than when they are again assayed in the wild, and as compared to separate individuals assayed in the wild, potentially because captivity is an unfamiliar situation.

P6 alternative 1: Individuals in captivity are more exploratory and bold than wild individuals (testing sessions matched for season), and captive individuals show more exploratory and bold behaviors than when they are subsequently tested in the wild, potentially because the captive environment decreases the influence of predation, social interactions and competition.

P6 alternative 2: There is no difference in exploration and boldness between individuals in captivity and individuals in the wild (matched for season), potentially because in both contexts our data is biased by sampling only the types of individuals that were most likely to get caught in traps.

P6 alternative 3: Captive individuals, when tested again after being released, show no difference in exploratory and bold behaviors because our methods assess inherent personality traits that are consistent across the captive and wild contexts in this taxa.

D. METHODS

Planned Sample

Great-tailed grackles are caught in the wild in Tempe, Arizona USA for individual identification (colored leg bands in unique combinations). Some individuals (~32) are brought temporarily into aviaries for testing, and then they will be released back to the wild. Grackles are individually housed in an aviary (each 244cm long by 122cm wide by 213cm tall) at Arizona State University for a maximum of three months where they have ad lib access to water at all times and are fed Mazuri Small Bird maintenance diet ad lib during non-testing hours (minimum 20h per day), and various other food items (e.g., peanuts, grapes, bread) during testing (up to 3h per day per bird). Individuals are given three to four days to habituate to the aviaries and then their test battery begins on the fourth or fifth day (birds are usually tested six days per week, therefore if their fourth day in the aviaries occurs on a day off, then they are tested on the fifth day instead). For hypothesis 2 we will attempt to test all grackles in the wild that are color-banded.

Sample size rationale

We will test as many birds as we can in the approximately three years at this field site given that the birds only participate in tests in aviaries during the non-breeding season (approximately September through March). The minimum sample size for captive subjects will be 16, however we expect to be able to test up to 32 grackles in captivity. We catch grackles with a variety of methods, some of which decrease the likelihood of a selection bias for exploratory and bold individuals because grackles cannot see the traps (i.e. mist nets). In sampling all banded birds in the wild, we will therefore have a better idea of the variation in exploration and boldness behaviors in this population.

Data collection stopping rule

We will stop testing birds once we have completed two full aviary seasons (likely in March 2020) if the sample size is above the minimum suggested boundary based on model simulations (see section “Ability to detect actual effects” below). If the minimum sample size is not met by this point, we will continue testing birds at our next field site (which we move to in the summer of 2020) until we meet the minimum sample size.

Open materials

Testing protocols (<https://docs.google.com/document/d/1sEMc5z2fw6S9C-wVfc2zV331CRPpu3NuA7lhSFUZJpE/edit?usp=sharing>) for exploration of new environments and objects, boldness, persistence, and motor diversity.

Open data

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

Randomization and counterbalancing

There is no randomizing. The order of the three tasks will be counterbalanced across birds (using <https://www.random.org> (<https://www.random.org>) to randomly assign individuals to one of three experimental orders).

1/3 of the individuals will experience:

1. Exploration environment
2. Exploration object
3. Boldness

1/3 of the individuals will experience:

1. Exploration object
2. Boldness
3. Exploration environment

1/3 of the individuals will experience:

1. Boldness
2. Exploration environment
3. Exploration object

Blinding of conditions during analysis

No blinding is involved in this study.

####Variables included in analyses 1-5

NOTE: to view a list of these variables in a table format, please see our Google sheet

(<https://docs.google.com/spreadsheets/d/1nhFkqTFWeAeWli8FU8n7mDiWGBuCeduzf8tWN3wPQeE/edit?usp=sharing>), which describes whether they are a dependent variable (DV), independent variable (IV), or random effect (RE). Note: when there is more than one DV per model, all models will be run once per DV.

ANALYSIS 1 - REPEATABILITY of boldness, persistence and exploration

Dependent variables

1. Boldness: Latency to land on the table - OR - Latency to eat the food - OR - Latency to touch a threatening object next to food (we will choose the variable with the most data)
2. Persistence: Number of touches to an apparatus per time (multi-access box in the behavioral flexibility preregistration (https://github.com/corinalogan/grackles/blob/master/EasyToReadFiles/g_flexmanip.md), novel environment in P1, and objects in P2 and P4)
3. Exploration of novel environment: Latency to enter a novel environment set inside a familiar environment
4. Exploration of novel object: Latency to land on the table next to an object (novel, familiar) (that does not contain food) in a familiar environment (that contains maintenance diet away from the object) - OR - latency to touch an object (novel, familiar) (choose the variable with the most data)

Independent variables

1. Condition (https://github.com/corinalogan/grackles/blob/master/EasyToReadFiles/g_flexmanip.md): control, flexibility manipulation

2. ID (random effect because multiple measures per individual)

ANALYSIS 2 - P1-P5: flexibility correlates with exploratory behaviors

Dependent variables

1. The **number of trials to reverse** a preference in the last reversal that individual participated in (an individual is considered to have a preference if it chose the rewarded option at least 17 out of the most recent 20 trials (with a minimum of 8 or 9 correct choices out of 10 on the two most recent sets of 10 trials)). See behavioral flexibility preregistration (https://github.com/corinalogan/grackles/blob/master/EasyToReadFiles/g_flexmanip.md) for details.
2. The **ratio of correct divided by incorrect trials** for the first 40 trials in their final reversal after the individual has seen the newly rewarded option once. These 40 trials include trials where individuals were offered the test and chose not to participate (i.e., make a choice). This accounts for flexibility that can occur when some individuals inhibit their previously rewarded preference (thus exhibiting flexibility because they changed their behavior when circumstances changed), but are not as exploratory as those who have fewer 'no choice' trials. 'No choice' data is data that is otherwise excluded from standard reversal learning analyses. Including 'no choice' trials, controls for individual differences in exploration because those that refuse to choose are not exploring new options, which would allow them to learn the new food location.
3. If the number of trials to reverse a preference does not positively correlate with the number of trials to attempt or solve new loci on the multi-access box (an additional measure of behavioral flexibility), then the **average number of trials to solve** and the **average number of trials to attempt** a new option on the multi-access box will be additional dependent variables. See behavioral flexibility preregistration (https://github.com/corinalogan/grackles/blob/master/EasyToReadFiles/g_flexmanip.md).
4. **Flexibility comprehensive:** This measure is currently being developed and is intended be a more accurate representation of all of the choices an individual made, as well as accounting for the degree of uncertainty exhibited by individuals as preferences change. If this measure more effectively represents flexibility (determined using a modeled dataset and not the actual data), we may decide to solely rely on this measure and not use independent variables 1-3. If this ends up being the case, we will modify the code in the analysis plan below to reflect this change before conducting analyses of the data in this preregistration.

All models will be run once per dependent variable.

Independent variables

1. P1: Latency to enter a novel environment inside a familiar environment
2. P1: 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
3. P1: Time spent per section of a novel environment or in the corresponding areas on othe 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
4. P1: Time spent exploring the outside of the novel environment (within 20cm) before entering it
5. P2: Latency to land on the table next to an object (novel, familiar) (that does not contain food) in a familiar environment (that contains maintenance diet away from the object) - OR - latency to touch an object (novel, familiar) (choose the variable with the most data)
6. P3: Number of touches to the functional part of an apparatus per time (multi-access box, novel environment in P1, novel objects in P2 and P4)
7. P3: Number of touches to the non-functional part of an apparatus per time (multi-access box (https://github.com/corinalogan/grackles/blob/master/EasyToReadFiles/g_flexmanip.md))
8. P4: Latency to land on the table - OR - Latency to eat the food - OR - Latency to touch a threatening object next to food (choose the variable with the most data)

9. P5: Number of different motor actions used when attempting to solve the multi-access box (https://github.com/corinalogan/grackles/blob/master/EasyToReadFiles/g_flexmanip.md)
10. Age (adult: after hatch year, juvenile: hatch year). NOTE: this variable will be removed if only adults are tested (and we are planning to test only adults).
11. ID (random effect because multiple measures per individual)
12. Condition (https://github.com/corinalogan/grackles/blob/master/EasyToReadFiles/g_flexmanip.md): control, flexibility manipulation

ANALYSIS 3 - P1 alternative 4: correlation between boldness and exploration

Dependent variable: Boldness: Latency to land on the table - OR - Latency to eat the food - OR - Latency to touch a threatening object next to food (we will choose the variable with the most data)

Independent variables:

1. Time spent exploring the outside of the novel environment (within 20cm) before entering it
2. Latency to land on the table next to an object (novel, familiar) (that does not contain food) in a familiar environment (that contains maintenance diet away from the object) - OR - latency to touch an object (novel, familiar) (choose the variable with the most data)

ANALYSIS 4 - P3: does persistence correlate with reversal persistence?

Dependent variable: The number of incorrect choices in the final reversal before making the first correct choice

Independent variables:

1. Average number of touches to the functional part of an apparatus per time (multi-access box (https://github.com/corinalogan/grackles/blob/master/EasyToReadFiles/g_flexmanip.md), novel environment in P1, novel objects in P2 and P4)
2. Condition: control, flexibility manipulation

ANALYSIS 5 - P6: captive vs wild

Dependent variables

1. Boldness: In captivity we will measure boldness as the latency to land on the table - OR - Latency to eat the food - OR - Latency to touch a threatening object that is next to food (we will choose the variable with the most data); In the wild the dependent variable will be the latency to come within 2m - OR - Latency to eat the food - OR - Latency to touch a threatening object that is next to food (we will choose the variable with the most data).
2. Persistence: Number of touches to an apparatus per time (multi-access box in the behavioral flexibility preregistration (https://github.com/corinalogan/grackles/blob/master/EasyToReadFiles/g_flexmanip.md), novel environment in P1, objects in P2 and P4)
3. Exploration of novel environment: Latency to enter a novel sub-environment inside a familiar environment
4. Exploration of novel object: Latency to land next to an object (novel, familiar) (that does not contain food) in a familiar environment (that contains maintenance diet away from the object) - OR - latency to touch an object (novel, familiar) (choose the variable with the most data)

Note: if 3 and 4 are consistent within individuals, and correlate, we will combine these variables into one exploration propensity score.

Independent variables

1. Context: captive or wild
2. Number of times we attempted to assay boldness or exploration but failed due to lack of participation
3. ID (random effect because multiple measures per individual)

E. ANALYSIS PLAN

We do not plan to **exclude** any data. When **missing data** occur, the existing data for that individual will be included in the analyses for the tests they completed. Analyses will be conducted in R (current version 4.0.3; (R Core Team 2017)). When there is more than one experimenter within a test, experimenter will be added as a random effect to account for potential differences between experimenters in conducting the tests. If there are no differences between models including or excluding experimenter as a random effect, then we will use the model without this random effect for simplicity.

Ability to detect actual effects

To begin to understand what kinds of effect sizes we will be able to detect given our sample size limitations and our interest in decreasing noise by attempting to measure it, which increases the number of explanatory variables, we used G*Power (v.3.1, Faul et al. (2007), Faul et al. (2009)) to conduct power analyses based on confidence intervals. G*Power uses pre-set drop down menus and we chose the options that were as close to our analysis methods as possible (listed in each analysis below). Note that there were no explicit options for GLMs (though the chosen test in G*Power appears to align with GLMs) or GLMMs or for the inclusion of the number of trials per bird (which are generally large in our investigation), thus the power analyses are only an approximation of the kinds of effect sizes we can detect. We realize that these power analyses are not fully aligned with our study design and that these kinds of analyses are not appropriate for Bayesian statistics (e.g., our MCMCglmm below), however we are unaware of better options at this time. Additionally, it is difficult to run power analyses because it is unclear what kinds of effect sizes we should expect due to the lack of data on this species for these experiments.

To address the power analysis issues, we will run simulations on our Arizona data set before conducting any analyses in this preregistration. We will first run null models (i.e., dependent variable ~ 1 + random effects), which will allow us to determine what a weak versus a strong effect is for each model. Then we will run simulations based on the null model to explore the boundaries of influences (e.g., sample size) on our ability to detect effects of interest of varying strengths. If simulation results indicate that our Arizona sample size is not larger than the lower boundary, we will continue these experiments at the next field site until we meet the minimum suggested sample size.

Data checking

The data will be checked for overdispersion, underdispersion, zero-inflation, and heteroscedasticity with the DHARMa R package (Hartig 2019) following methods by Hartig (<https://cran.r-project.org/web/packages/DHARMa/vignettes/DHARMa.html>). Note: DHARMa doesn't support MCMCglmm, therefore we will use the closest supported model: glmer from the R package lme4 (Douglas Bates et al. 2015).

Repeatability of exploration, boldness and persistence

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

Note: The power analysis is the same as for P3 (below) because there are the same number of explanatory variables (fixed effects).

```
# Boldness
bol <- read.csv("/Users/corina/GTGR/data/data_boldness.csv",
  header = T, sep = ",", stringsAsFactors = F)

# DATA CHECKING
library(DHARMa)
library(lme4)
simulationOutput <- simulateResiduals(fittedModel = glmer(LatencyToFeed ~
  Condition + (1 | ID), family = poisson, data = bol), n = 250) #250 simulations, but if want higher p
recision change n>1000
simulationOutput$scaledResiduals #Expect a flat distribution of the overall residuals, and uniformity in
```



```

y direction if plotted against any predictor
testDispersion(simulationOutput) #if under- or over-dispersed, then p-value<0.05, but then check the dispersion parameter and try to determine what in the model could be the cause and address it there, also check for zero inflation
testZeroInflation(simulationOutput) #compare expected vs observed zeros, not zero-inflated if p<0.05
testUniformity(simulationOutput) #check for heteroscedasticity ('a systematic dependency of the dispersion / variance on another variable in the model' Hartig, https://cran.r-project.org/web/packages/DHARMA/vignettes/DHARMA.html), which is indicated if dots aren't on the red line and p<0.05. Also...
plot(simulationOutput) #...there should be no pattern in the data points in the right panel
plotResiduals(ReverseNumber, simulationOutput$scaledResiduals) #plot the residuals against other predictors (in cases when there is more than 1 fixed effect) - can't get this code to work yet

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

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

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

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

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

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

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

```

```

# Persistence
per <- read.csv("/Users/corina/GTGR/data/data_persist.csv", header = T,
  sep = ",", stringsAsFactors = F)

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

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

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

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

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

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

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

```

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

```
# Exploration of novel environment
ee <- read.csv("/Users/corina/GTGR/data/data_explore.csv", header = T,
  sep = ",", stringsAsFactors = F)

# DATA CHECKING
library(DHARMA)
library(lme4)
simulationOutput <- simulateResiduals(fittedModel = glmer(LatencyExpEnv ~
  Condition + (1 | ID), family = poisson, data = ee), n = 250) #250 simulations, but if want higher pr
ecision change n>1000
simulationOutput$scaledResiduals #Expect a flat distribution of the overall residuals, and uniformity in
y direction if plotted against any predictor
testDispersion(simulationOutput) #if under- or over-dispersed, then p-value<0.05, but then check the dis
persion parameter and try to determine what in the model could be the cause and address it there, also ch
eck for zero inflation
testZeroInflation(simulationOutput) #compare expected vs observed zeros, not zero-inflated if p<0.05
testUniformity(simulationOutput) #check for heteroscedasticity ('a systematic dependency of the dispersi
on / variance on another variable in the model' Hartig, https://cran.r-project.org/web/packages/DHARMA/vi
gnettes/DHARMA.html), which is indicated if dots aren't on the red line and p<0.05. Also...
plot(simulationOutput) #...there should be no pattern in the data points in the right panel
plotResiduals(ReverseNumber, simulationOutput$scaledResiduals) #plot the residuals against other predict
ors (in cases when there is more than 1 fixed effect) - can't get this code to work yet

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

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

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

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

```

})) #estimates for each iteration of the MCMC

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

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

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

```

```

# Exploration of novel object
eo <- read.csv("/Users/corina/GTGR/data/data_persist.csv", header = T,
  sep = ",", stringsAsFactors = F)

# DATA CHECKING
library(DHARMA)
library(lme4)
simulationOutput <- simulateResiduals(fittedModel = glmer(LatencyTableExpObject ~
  Condition + (1 | ID), family = poisson, data = eo), n = 250) #250 simulations, but if want higher pr
ecision change n>1000
simulationOutput$scaledResiduals #Expect a flat distribution of the overall residuals, and uniformity in
y direction if plotted against any predictor
testDispersion(simulationOutput) #if under- or over-dispersed, then p-value<0.05, but then check the dis
persion parameter and try to determine what in the model could be the cause and address it there, also ch
eck for zero inflation
testZeroInflation(simulationOutput) #compare expected vs observed zeros, not zero-inflated if p<0.05
testUniformity(simulationOutput) #check for heteroscedasticity ('a systematic dependency of the dispersi
on / variance on another variable in the model' Hartig, https://cran.r-project.org/web/packages/DHARMA/vi
gnettes/DHARMA.html), which is indicated if dots aren't on the red line and p<0.05. Also...
plot(simulationOutput) #...there should be no pattern in the data points in the right panel
plotResiduals(ReverseNumber, simulationOutput$scaledResiduals) #plot the residuals against other predict
ors (in cases when there is more than 1 fixed effect) - can't get this code to work yet

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

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

```

```

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

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

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

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

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

```

P1-P5: correlation of flexibility with exploration of new environments and objects, boldness, persistence, and motor diversity

Analysis: If behavior is not repeatable across assays at Time 1 and Time 2 (six weeks apart, both assays occur after the flexibility manipulation takes place) for exploration, boldness, persistence, or motor diversity (see analysis for P6), we will not include these variables in analyses involving flexibility. If behavior is repeatable within individuals, we will examine the relationship between flexibility and these variables as follows. Note that the two exploration measures (novel environment and novel object) will be combined into one variable if they correlate and are both repeatable within individuals.

Because the independent variables could influence each other, we will analyze them in a single model: Generalized Linear Mixed Model (GLMM; MCMCglmm function, MCMCglmm package; (J. D. Hadfield 2010)) with a Poisson distribution and log link using 13,000 iterations with a thinning interval of 10, a burnin of 3,000, and minimal priors ($V=1$, $\nu=0$) (J. Hadfield 2014). We will ensure the GLMM shows acceptable convergence (i.e., lag time autocorrelation values <0.01 ; (J. D. Hadfield 2010)), and adjust parameters if necessary. We will determine whether an independent variable had an effect or not using the Estimate in the full model.

To roughly estimate our ability to detect actual effects (because these power analyses are designed for frequentist statistics, not Bayesian statistics), we ran a power analysis in G*Power with the following settings: test family=F tests, statistical test=linear multiple regression: Fixed model (R^2 deviation from zero), type of power analysis=a priori, alpha error probability=0.05. We reduced the power to 0.70 and increased the effect size until the total sample size in the output matched our projected sample size ($n=32$). The number of predictor variables was restricted to only the fixed effects because this test was not designed for mixed models. The protocol of the power analysis is here:

Input:

Effect size $f^2 = 0.62$

α err prob = 0,05

Power ($1-\beta$ err prob - note: β =probability of making a Type II error) = 0,7

Number of predictors = 10

Output:

Noncentrality parameter λ = 19,8400000

Critical F = 2,3209534

Numerator df = 10

Denominator df = 21

Total sample size = 32

Actual power = 0,7027626

This means that, with our sample size of 32, we have a 70% chance of detecting a large effect (approximated at $f^2=0.35$ by Cohen (1988)).

```
explore <- read.csv("/Users/corina/GTGR/data/data_explore.csv",
  header = T, sep = ",", stringsAsFactors = F)

# DATA CHECKING for 1st GLMM
library(DHARMA)
library(lme4)
simulationOutput <- simulateResiduals(fittedModel = glmer(TrialsToReverseLast ~
  Condition + TimeOutsideNovelEnv + LatencyExpEnv + AverageTimePerSection *
  EnvCondition + TotalNumberSections * EnvCondition + LatencyTableExpObject +
  MultiaccessTouchesPerTime + LatencyBoldness + NoMotorActions +
  (1 | Batch), family = poisson, data = explore), n = 250) #250 simulations, but if want higher pr
ecision change n>1000
simulationOutput$scaledResiduals #Expect a flat distribution of the overall residuals, and uniformity in
y direction if plotted against any predictor
testDispersion(simulationOutput) #if under- or over-dispersed, then p-value<0.05, but then check the dis
persion parameter and try to determine what in the model could be the cause and address it there, also ch
eck for zero inflation
testZeroInflation(simulationOutput) #compare expected vs observed zeros, not zero-inflated if p<0.05
testUniformity(simulationOutput) #check for heteroscedasticity ('a systematic dependency of the dispersi
on / variance on another variable in the model' Hartig, https://cran.r-project.org/web/packages/DHARMA/vi
gnettes/DHARMA.html), which is indicated if dots aren't on the red line and p<0.05. Also...
plot(simulationOutput) #...there should be no pattern in the data points in the right panel
plotResiduals(ReverseNumber, simulationOutput$scaledResiduals) #plot the residuals against other predict
ors (in cases when there is more than 1 fixed effect) - can't get this code to work yet

# ANALYSIS Take the average of Time 1 and Time 2 for each
# variable (exploration environment, exploration object,
# boldness, motor diversity, persistence)
TimeOutsideNovelEnvT1 <- TimeOutsideNovelEnv[TimeOutsideNovelEnv$Time ==
  1, ]
TimeOutsideNovelEnvT2 <- TimeOutsideNovelEnv[TimeOutsideNovelEnv$Time ==
  2, ]
TimeOutsideNovelEnv <- (TimeOutsideNovelEnvT1 + TimeOutsideNovelEnvT2)/2

LatencyExpEnvT1 <- LatencyExpEnv[LatencyExpEnv$Time == 1, ]
LatencyExpEnvT2 <- LatencyExpEnv[LatencyExpEnv$Time == 2, ]
LatencyExpEnv <- (LatencyExpEnvT1 + LatencyExpEnvT2)/2

AverageTimePerSectionNovelEnvT1 <- AverageTimePerSectionNovelEnv[AverageTimePerSectionNovelEnv$Time ==
```

```

1, ]
AverageTimePerSectionNovelEnvT2 <- AverageTimePerSectionNovelEnv[AverageTimePerSectionNovelEnv$Time ==
2, ]
AverageTimePerSectionNovelEnv <- (AverageTimePerSectionNovelEnvT1 +
  AverageTimePerSectionNovelEnvT2)/2

TotalNumberSectionsNovelEnvT1 <- TotalNumberSectionsNovelEnv[TotalNumberSectionsNovelEnv$Time ==
1, ]
TotalNumberSectionsNovelEnvT2 <- TotalNumberSectionsNovelEnv[TotalNumberSectionsNovelEnv$Time ==
2, ]
TotalNumberSectionsNovelEnv <- (TotalNumberSectionsNovelEnvT1 +
  TotalNumberSectionsNovelEnvT2)/2

LatencyTableExpObjectT1 <- LatencyTableExpObject[LatencyTableExpObject$Time ==
1, ]
LatencyTableExpObjectT2 <- LatencyTableExpObject[LatencyTableExpObject$Time ==
2, ]
LatencyTableExpObject <- (LatencyTableExpObjectT1 + LatencyTableExpObjectT2)/2

MultiaccessTouchesPerTimeT1 <- MultiaccessTouchesPerTime[MultiaccessTouchesPerTime$Time ==
1, ]
MultiaccessTouchesPerTimeT2 <- MultiaccessTouchesPerTime[MultiaccessTouchesPerTime$Time ==
2, ]
MultiaccessTouchesPerTime <- (MultiaccessTouchesPerTimeT1 + MultiaccessTouchesPerTimeT2)/2

LatencyBoldnessT1 <- LatencyBoldness[LatencyBoldness$Time ==
1, ]
LatencyBoldnessT2 <- LatencyBoldness[LatencyBoldness$Time ==
2, ]
LatencyBoldness <- (LatencyBoldnessT1 + LatencyBoldnessT2)/2

NoMotorActionsT1 <- NoMotorActions[NoMotorActions$Time == 1,
]
NoMotorActionsT1 <- NoMotorActions[NoMotorActions$Time == 1,
]
NoMotorActions <- (NoMotorActionsT1 + NoMotorActionsT2)/2

# GLMM - dependent variable = number of trials to reverse
# in the last reversal
library(MCMCglmm)
prior = list(R = list(R1 = list(V = 1, nu = 0), R2 = list(V = 1,
  nu = 0), R3 = list(V = 1, nu = 0), R4 = list(V = 1, nu = 0),
  R5 = list(V = 1, nu = 0), R6 = list(V = 1, nu = 0), R7 = list(V = 1,
    nu = 0), R8 = list(V = 1, nu = 0), R9 = list(V = 1, nu = 0),
    R10 = list(V = 1, nu = 0), R11 = list(V = 1, nu = 0)), G = list(G1 = list(V = 1,
    nu = 0), G2 = list(V = 1, nu = 0)))

expl1 <- MCMCglmm(TrialsToReverseLast ~ Condition + TimeOutsideNovelEnv +
  LatencyExpEnv + AverageTimePerSection * EnvCondition + TotalNumberSections *
  EnvCondition + LatencyTableExpObject + MultiaccessTouchesPerTime +
  LatencyBoldness + NoMotorActions, random = ~Batch, family = "poisson",
  data = explore, verbose = F, prior = prior, nitt = 13000,
  thin = 10, burnin = 3000)
summary(expl1)
autocorr(expl1$Sol) #Did fixed effects converge?
autocorr(expl1$VCV) #Did random effects converge?

```

```

# DATA CHECKING for 2nd GLMM
library(DHARMA)
library(lme4)
simulationOutput <- simulateResiduals(fittedModel = glmer(Ratio40 ~
  Condition + TimeOutsideNovelEnv + LatencyExpEnv + AverageTimePerSection *
  EnvCondition + TotalNumberSections * EnvCondition + LatencyTableExpObject +
  MultiaccessTouchesPerTime + LatencyBoldness + NoMotorActions +
  (1 | Batch), family = poisson, data = explore), n = 250) #250 simulations, but if want higher pr
ecision change n>1000
simulationOutput$scaledResiduals #Expect a flat distribution of the overall residuals, and uniformity in
y direction if plotted against any predictor
testDispersion(simulationOutput) #if under- or over-dispersed, then p-value<0.05, but then check the dis
persion parameter and try to determine what in the model could be the cause and address it there, also ch
eck for zero inflation
testZeroInflation(simulationOutput) #compare expected vs observed zeros, not zero-inflated if p<0.05
testUniformity(simulationOutput) #check for heteroscedasticity ('a systematic dependency of the dispersi
on / variance on another variable in the model' Hartig, https://cran.r-project.org/web/packages/DHARMA/vi
gnettes/DHARMA.html), which is indicated if dots aren't on the red line and p<0.05. Also...
plot(simulationOutput) #...there should be no pattern in the data points in the right panel
plotResiduals(ReverseNumber, simulationOutput$scaledResiduals) #plot the residuals against other predict
ors (in cases when there is more than 1 fixed effect) - can't get this code to work yet

# 2nd GLMM - dependent variable = ratio of correct divided
# by incorrect trials for first 40 trials of final reversal
# after making the first correct choice

expl2 <- MCMCglmm(Ratio40 ~ Condition + TimeOutsideNovelEnv +
  LatencyExpEnv + AverageTimePerSection * EnvCondition + TotalNumberSections *
  EnvCondition + LatencyTableExpObject + MultiaccessTouchesPerTime +
  LatencyBoldness + NoMotorActions, random = ~Batch, family = "poisson",
  data = explore, verbose = F, prior = prior, nitt = 13000,
  thin = 10, burnin = 3000)

summary(expl2)
autocorr(expl2$Sol) #Did fixed effects converge?
autocorr(expl2$VCV) #Did random effects converge?

```

P1 alternative 4: correlations between exploration and boldness measures

Analysis: Generalized Linear Model (GLM; glm function, stats package) with a Poisson distribution and log link. For an estimation of our ability to detect actual effects, please see the power analysis for P3 below.


```

expbol <- read.csv("/Users/corina/GTGR/data/data_reverselatency.csv",
  header = T, sep = ",", stringsAsFactors = F)

# DATA CHECKING
library(DHARMA)
library(lme4)
simulationOutput <- simulateResiduals(fittedModel = glmer(LatencyBoldness ~
  TimeOutsideNovelEnv + LatencyTableExpObject, family = poisson,
  data = persist2), n = 250) #250 simulations, but if want higher precision change n>1000
simulationOutput$scaledResiduals #Expect a flat distribution of the overall residuals, and uniformity in
y direction if plotted against any predictor
testDispersion(simulationOutput) #if under- or over-dispersed, then p-value<0.05, but then check the dis
persion parameter and try to determine what in the model could be the cause and address it there, also ch
eck for zero inflation
testZeroInflation(simulationOutput) #compare expected vs observed zeros, not zero-inflated if p<0.05
testUniformity(simulationOutput) #check for heteroscedasticity ('a systematic dependency of the dispersi
on / variance on another variable in the model' Hartig, https://cran.r-project.org/web/packages/DHARMA/vi
gnettes/DHARMA.html), which is indicated if dots aren't on the red line and p<0.05. Also...
plot(simulationOutput) #...there should be no pattern in the data points in the right panel
plotResiduals(ReverseNumber, simulationOutput$scaledResiduals) #plot the residuals against other predict
ors (in cases when there is more than 1 fixed effect) - can't get this code to work yet

# GLM
eb <- glm(LatencyBoldness ~ TimeOutsideNovelEnv + LatencyTableExpObject,
  family = "poisson", data = expbol)
# summary(eb)

eb2 <- summary(eb)
library(xtable)
eb2.table <- xtable(eb2)
library(knitr)
kable(eb2.table, caption = "Table 2: Model selection output.",
  format = "html", digits = 2)

# Model Validation
library(MuMIn)
options(na.action = "na.fail")
basel <- dredge(glm(NumberIncorrectTrialsReversal ~ AvgFunctionalTouchesExploration *
  Condition, family = "poisson", data = persist2))
library(knitr)
kable(basel, caption = "Table 3: Model selection output.")

```

Model validation: Determine whether the test model results are likely to be reliable given the data (Burnham and Anderson 2003). Compare Akaike weights (range: 0–1, the sum of all model weights equals 1; Akaike, 1981) between the test model and a base model (number of trials to reverse as the response variable and 1 as the explanatory variable) using the dredge function in the MuMIn package (D. Bates, Maechler, and Bolker 2012). If the best fitting model has a high Akaike weight (>0.89; (Burnham and Anderson 2003)), then it indicates that the results are likely given the data. The Akaike weights indicate the best fitting model is the [base/test - delete as appropriate] model (Table 2).

P3: correlations between persistence measures

Analysis: Generalized Linear Model (GLM; glm function, stats package) with a Poisson distribution and log link.

To determine our ability to detect actual effects, we ran a power analysis in G*Power with the following settings: test family=F tests, statistical test=linear multiple regression: Fixed model (R^2 deviation from zero), type of power analysis=a priori, alpha error probability=0.05. We reduced the power to 0.70 and increased the effect size until the total sample size in the output matched our projected sample size (n=32). The protocol of the power analysis is here:

Input:

Effect size $f^2 = 0,27$

α err prob = 0,05

Power ($1 - \beta$ err prob - note: β = probability of making a Type II error) = 0,7

Number of predictors = 2

Output:

Noncentrality parameter $\lambda = 8,6400000$

Critical F = 3,3276545

Numerator df = 2

Denominator df = 29

Total sample size = 32

Actual power = 0,7047420

This means that, with our sample size of 32, we have a 70% chance of detecting a medium (approximated at $f^2=0.15$ by Cohen (1988)) to large effect (approximated at $f^2=0.35$ by Cohen (1988)).

```

persist2 <- read.csv("/Users/corina/GTGR/data/data_reverselatency.csv",
  header = T, sep = ",", stringsAsFactors = F)

# DATA CHECKING
library(DHARMA)
library(lme4)
simulationOutput <- simulateResiduals(fittedModel = glmer(NumberIncorrectTrialsReversal ~
  AvgFunctionalTouchesExploration * Condition, family = poisson,
  data = persist2), n = 250) #250 simulations, but if want higher precision change n>1000
simulationOutput$scaledResiduals #Expect a flat distribution of the overall residuals, and uniformity in
y direction if plotted against any predictor
testDispersion(simulationOutput) #if under- or over-dispersed, then p-value<0.05, but then check the dis
persion parameter and try to determine what in the model could be the cause and address it there, also ch
eck for zero inflation
testZeroInflation(simulationOutput) #compare expected vs observed zeros, not zero-inflated if p<0.05
testUniformity(simulationOutput) #check for heteroscedasticity ('a systematic dependency of the dispersi
on / variance on another variable in the model' Hartig, https://cran.r-project.org/web/packages/DHARMA/vi
gnettes/DHARMA.html), which is indicated if dots aren't on the red line and p<0.05. Also...
plot(simulationOutput) #...there should be no pattern in the data points in the right panel
plotResiduals(ReverseNumber, simulationOutput$scaledResiduals) #plot the residuals against other predict
ors (in cases when there is more than 1 fixed effect) - can't get this code to work yet

# GLM
p2 <- glm(NumberIncorrectTrialsReversal ~ AvgFunctionalTouchesExploration *
  Condition, family = "poisson", data = persist2)
# summary(p2)

sp2 <- summary(p2)
library(xtable)
sp2.table <- xtable(sp2)
library(knitr)
kable(sp2.table, caption = "Table 2: Model selection output.",
  format = "html", digits = 2)

# Model Validation
library(MuMIn)
options(na.action = "na.fail")
basel <- dredge(glm(NumberIncorrectTrialsReversal ~ AvgFunctionalTouchesExploration *
  Condition, family = "poisson", data = persist2))
library(knitr)
kable(basel, caption = "Table 3: Model selection output.")

```

Model validation: Determine whether the test model results are likely to be reliable given the data (Burnham and Anderson 2003). Compare Akaike weights (range: 0–1, the sum of all model weights equals 1; Akaike, 1981) between the test model and a base model (number of trials to reverse as the response variable and 1 as the explanatory variable) using the dredge function in the MuMIn package (D. Bates, Maechler, and Bolker 2012). If the best fitting model has a high Akaike weight (>0.89; (Burnham and Anderson 2003)), then it indicates that the results are likely given the data. The Akaike weights indicate the best fitting model is the [base/test - delete as appropriate] model (Table 2).

Alternative Analyses

We anticipate that we will want to run additional/different analyses after reading McElreath (2016). We will revise this preregistration to include these new analyses before conducting the analyses above.

F. ETHICS

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

1. US Fish and Wildlife Service (scientific collecting permit number MB76700A-0,1,2)
2. US Geological Survey Bird Banding Laboratory (federal bird banding permit number 23872)
3. Arizona Game and Fish Department (scientific collecting license number SP594338 [2017] and SP606267 [2018])
4. Institutional Animal Care and Use Committee at Arizona State University (protocol number 17-1594R)
5. University of Cambridge ethical review process (non-regulated use of animals in scientific procedures: zoo4/17)

G. AUTHOR CONTRIBUTIONS

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

Rowney: Data collection, data interpretation, revising/editing.

Bergeron: Data collection, data interpretation, revising/editing.

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

H. 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.

I. ACKNOWLEDGEMENTS

We thank Jeremy Van Cleve and two anonymous reviewers for their feedback on the preregistration; Dieter Lukas for help polishing the hypotheses and for assistance with responding to reviewer comments; Ben Trumble for providing us with a wet lab at Arizona State University; Melissa Wilson Sayres for sponsoring our affiliations at Arizona State University and lending lab equipment; Kristine Johnson for technical advice on great-tailed grackles; Jay Taylor for grackle scouting at Arizona State University; Arizona State University School of Life Sciences Department Animal Care and Technologies for providing space for our aviaries and for their excellent support of our daily activities; Julia Cissewski for tirelessly solving problems involving financial transactions and contracts; Richard McElreath for project support; 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.

J. REFERENCES

- Bates, D, M Maechler, and B Bolker. 2012. "lme4: Linear Mixed-Effects Models Using Eigen and S4" (2011). R Package Version 0.999375-42.
- Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. "Fitting Linear Mixed-Effects Models Using lme4." *Journal of Statistical Software* 67 (1): 1–48. <https://doi.org/10.18637/jss.v067.i01> (<https://doi.org/10.18637/jss.v067.i01>).
- Burnham, Kenneth P, and David R Anderson. 2003. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer Science & Business Media.
- Chow, Pizza Ka Yee, Stephen EG Lea, and Lisa A Leaver. 2016. "How Practice Makes Perfect: The Role of Persistence, Flexibility and Learning in Problem-Solving Efficiency." *Animal Behaviour* 112: 273–83. <https://doi.org/10.1016/j.anbehav.2015.11.014> (<https://doi.org/10.1016/j.anbehav.2015.11.014>).
- Cohen, Jacob. 1988. "Statistical Power Analysis for the Behavioral Sciences 2nd Edn." Erlbaum Associates, Hillsdale.
- Faul, Franz, Edgar Erdfelder, Axel Buchner, and Albert-Georg Lang. 2009. "Statistical Power Analyses Using g* Power 3.1: Tests for Correlation and Regression Analyses." *Behavior Research Methods* 41 (4): 1149–60. <https://doi.org/10.3758/BRM.41.4.1149> (<https://doi.org/10.3758/BRM.41.4.1149>).
- Faul, Franz, Edgar Erdfelder, Albert-Georg Lang, and Axel Buchner. 2007. "G* Power 3: A Flexible Statistical Power Analysis Program for the Social, Behavioral, and Biomedical Sciences." *Behavior Research Methods* 39 (2): 175–91. <https://doi.org/10.3758/BF03193146> (<https://doi.org/10.3758/BF03193146>).
- Griffin, Andrea S, and David Guez. 2014. "Innovation and Problem Solving: A Review of Common Mechanisms." *Behavioural Processes* 109: 121–34. <https://doi.org/10.1016/j.beproc.2014.08.027> (<https://doi.org/10.1016/j.beproc.2014.08.027>).
- Hadfield, Jarrod D. 2010. "MCMC Methods for Multi-Response Generalized Linear Mixed Models: The MCMCglmm R Package."

- Journal of Statistical Software* 33 (2): 1–22. <http://www.jstatsoft.org/v33/i02/> (<http://www.jstatsoft.org/v33/i02/>).
- Hadfield, JD. 2014. “MCMCglmm Course Notes.” <http://cran.r-project.org/web/packages/MCMCglmm/vignettes/CourseNotes.pdf> (<http://cran.r-project.org/web/packages/MCMCglmm/vignettes/CourseNotes.pdf>).
- Hartig, Florian. 2019. *DHARMA: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models*. <http://florianhartig.github.io/DHARMA/> (<http://florianhartig.github.io/DHARMA/>).
- Lefebvre, Louis, Patrick Whittle, Evan Lascaris, and Adam Finkelstein. 1997. “Feeding Innovations and Forebrain Size in Birds.” *Animal Behaviour* 53 (3): 549–60. <https://doi.org/10.1006/anbe.1996.0330> (<https://doi.org/10.1006/anbe.1996.0330>).
- Logan, C J. 2016. “Behavioral Flexibility in an Invasive Bird Is Independent of Other Behaviors.” *PeerJ* 4: e2215.
- McElreath, Richard. 2016. *Statistical Rethinking: A Bayesian Course with Examples in r and Stan*. CRC Press. <https://doi.org/10.1201/9781315372495> (<https://doi.org/10.1201/9781315372495>).
- Mettke-Hofmann, Claudia, Sandy Lorentzen, Emmi Schlicht, Janine Schneider, and Franziska Werner. 2009. “Spatial Neophilia and Spatial Neophobia in Resident and Migratory Warblers (Sylvia).” *Ethology* 115 (5): 482–92.
- Mikhalevich, Irina, Russell Powell, and Corina Logan. 2017. “Is Behavioural Flexibility Evidence of Cognitive Complexity? How Evolution Can Inform Comparative Cognition.” *Interface Focus* 7 (3): 20160121. <https://doi.org/10.1098/rsfs.2016.0121> (<https://doi.org/10.1098/rsfs.2016.0121>).
- Peer, Brian D. 2011. “Invasion of the Emperor’s Grackle.” *Ardeola* 58 (2): 405–9.
- R Core Team. 2017. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org> (<https://www.R-project.org>).
- Sol, Daniel, Richard P Duncan, Tim M Blackburn, Phillip Cassey, and Louis Lefebvre. 2005. “Big Brains, Enhanced Cognition, and Response of Birds to Novel Environments.” *Proceedings of the National Academy of Sciences of the United States of America* 102 (15): 5460–65. <https://doi.org/10.1073/pnas.0408145102> (<https://doi.org/10.1073/pnas.0408145102>).
- Sol, Daniel, and Louis Lefebvre. 2000. “Behavioural Flexibility Predicts Invasion Success in Birds Introduced to New Zealand.” *Oikos* 90 (3): 599–605. <https://doi.org/10.1034/j.1600-0706.2000.900317.x> (<https://doi.org/10.1034/j.1600-0706.2000.900317.x>).
- Sol, Daniel, Sarah Timmermans, and Louis Lefebvre. 2002. “Behavioural Flexibility and Invasion Success in Birds.” *Animal Behaviour* 63 (3): 495–502.
- Wehtje, Walter. 2003. “The Range Expansion of the Great-Tailed Grackle (*Quiscalus Mexicanus* Gmelin) in North America Since 1880.” *Journal of Biogeography* 30 (10): 1593–1607. <https://doi.org/10.1046/j.1365-2699.2003.00970.x> (<https://doi.org/10.1046/j.1365-2699.2003.00970.x>).