Are the more flexible great-tailed grackles also better at inhibition?

Dr. Corina Logan (Max Planck Institute for Evolutionary Anthropology, corina_logan@eva.mpg.de), Dr. Kelsey McCune (University of California Santa Barbara / Max Planck Institute for Evolutionary Anthropology), Dr. Zoe Johnson-Ulrich (University of California Santa Barbara / Max Planck Institute for Evolutionary Anthropology), Luisa Bergeron (University of California Santa Barbara / Max Planck Institute for Evolutionary Anthropology), Carolyn Rowney (University of California Santa Barbara / Max Planck Institute for Evolutionary Anthropology), Benjamin Seitz (University of California Los Angeles), Dr. Aaron Blaisdell (University of California Los Angeles), Dr. Claudia Wascher (Anglia Ruskin University)

2018-10-17

```
#Make code wrap text so it doesn't go off the page when Knitting to PDF
library(knitr)
opts_chunk$set(tidy.opts=list(width.cutoff=60),tidy=TRUE)
```

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)) (Fig. 1). **This investigation**: In this piece of the long-term project, we aim to measure grackle inhibition in three experiments (delay of gratification, go-no go, detour) to determine whether those individuals that are more flexible are also better at inhibiting. Results will allow us to determine whether inhibition is linked with measures of flexibility (reversal learning and solution switching), how performance on these inhibition tests relate to each other to determine whether they measure the same or different traits, and validate the use of an inhibition task using a touch screen.

Figure 1. Five-year project overview. The same individuals will experience the experiments listed in each column (i.e., the same ~32 individuals in the left column (Years 1 and 2) will experience numbers 1-9, and the same ~32 individuals in the right column (Years 3-5) will experience A-D, plus numbers 5-9).

A. STATE OF THE DATA

Prior to collecting any data: This preregistration was written.

After data collection had begun: This preregistration was submitted to PCI Ecology for peer review after starting data collection on the detour task for the pre-reversal subcategory of subjects (for which there was data from one bird).

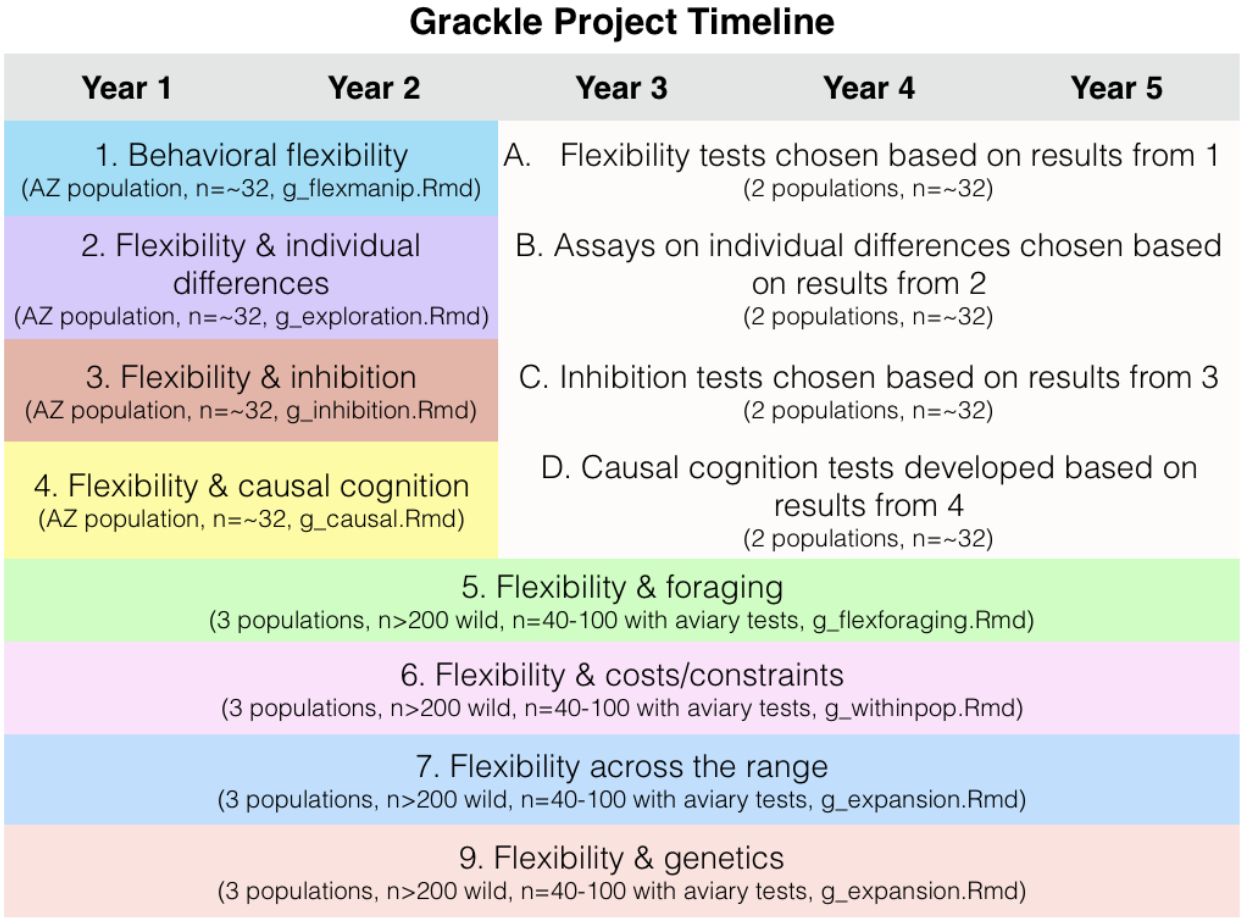


Figure 1: Figure 1. Five-year project overview.

B. PARTITIONING THE RESULTS

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

C. HYPOTHESIS

If flexibility requires inhibition, then individuals that are more behaviorally flexible (indicated by individuals that are faster at functionally changing their behavior when circumstances change), as measured by reversal learning and switching between options on a multi-access box, will also be better at inhibiting their responses in three tasks: delayed gratification, go no-go, and detour.

P1: Individuals that are faster to reverse preferences on a reversal learning task and who also have lower latencies to successfully solve new loci after previously solved loci become unavailable (multi-access box) (see flexibility preregistration) will perform better in the go no-go task (methods similar to Harding, Paul, and Mendl (2004)), in the detour task (methods as in MacLean et al. (2014) who call it the "cylinder task"), and they will wait longer for higher quality (more preferred) food, but not for higher quantities (methods as in Hillemann et al. (2014)). Waiting for higher quality food has been validated as a test of inhibition in birds, while waiting for a higher quantity of food does not appear to measure inhibition (Hillemann et al. (2014)).

P1 alternative 1: If there is no correlation between flexibility measures and performance on the inhibition tasks, this may indicate that the flexibility tasks may not require much inhibition (particularly if the inhibition results are reliable - see *P1 alternative 2*).

P1 alternative 2: If there is no correlation between flexibility measures and performance on the inhibition tasks, this may indicate that the inhibition tasks had low reliability and were therefore too noisy to correlate with flexibility.

P2: If there is no correlation in performance across inhibition tasks, it may indicate that that one or more of these tasks does not measure inhibition, or that they measure different types of inhibition (see Friedman and Miyake (2004)).

P2 alternative: If go no-go task performance strongly correlates with performance on the delayed gratification task, this indicates these two tasks measure the same trait, which therefore validates an inhibition task using a touch screen (the go no-go task).

P3: If individuals perform well on the detour task and with little individual variation, this is potentially because they will have had extensive experience looking into the sides of opaque tubes during reversal learning. To determine whether prior experience with opaque tubes in reversal learning contributed to their detour performance, a subset of individuals will experience the detour task before any reversal learning tests. If this subset performs the same as the others, then previous experience with tubes does not influence detour task performance. If the subset performs worse than the others, this indicates that detour task performance depends on the previous experiences of the indviduals tested.

Figure 2. The experimental designs of the three tasks: delayed gratification, go no-go, and detour (see protocol for details). In the **delay of gratification** task, individuals learn that food items will be transferred by the experimenter from a storing lid (near the experimenter) to a serving lid (near the bird) one at a time, and that they have access to the food in the serving lid from which they can eat at any time: they will learn that they will have access to more food if they wait longer for the experimenter to transfer food items. Once they pass training (by waiting for more than one food item in three trials), they move on to the test where food items are transferred from the serving to the storing lid with delays ranging from 2-1280 seconds. Birds will be tested on whether they are willing to wait for food items that increase in quality (i.e., are more preferred) or increase in quantity (i.e., the same food type accumulates in the serving lid). In the **go no-go** task, after pecking a start key on the touch screen to show they are interested in attending to a trial, they will see either a green circle or a purple circle (the rewarded circle color is counterbalanced across birds). Pecking the food key while the rewarded colored circle (green in the figure) is on the screen will result in the food hopper rising so the bird can eat food for 2 seconds, after which point the trial ends and the screen

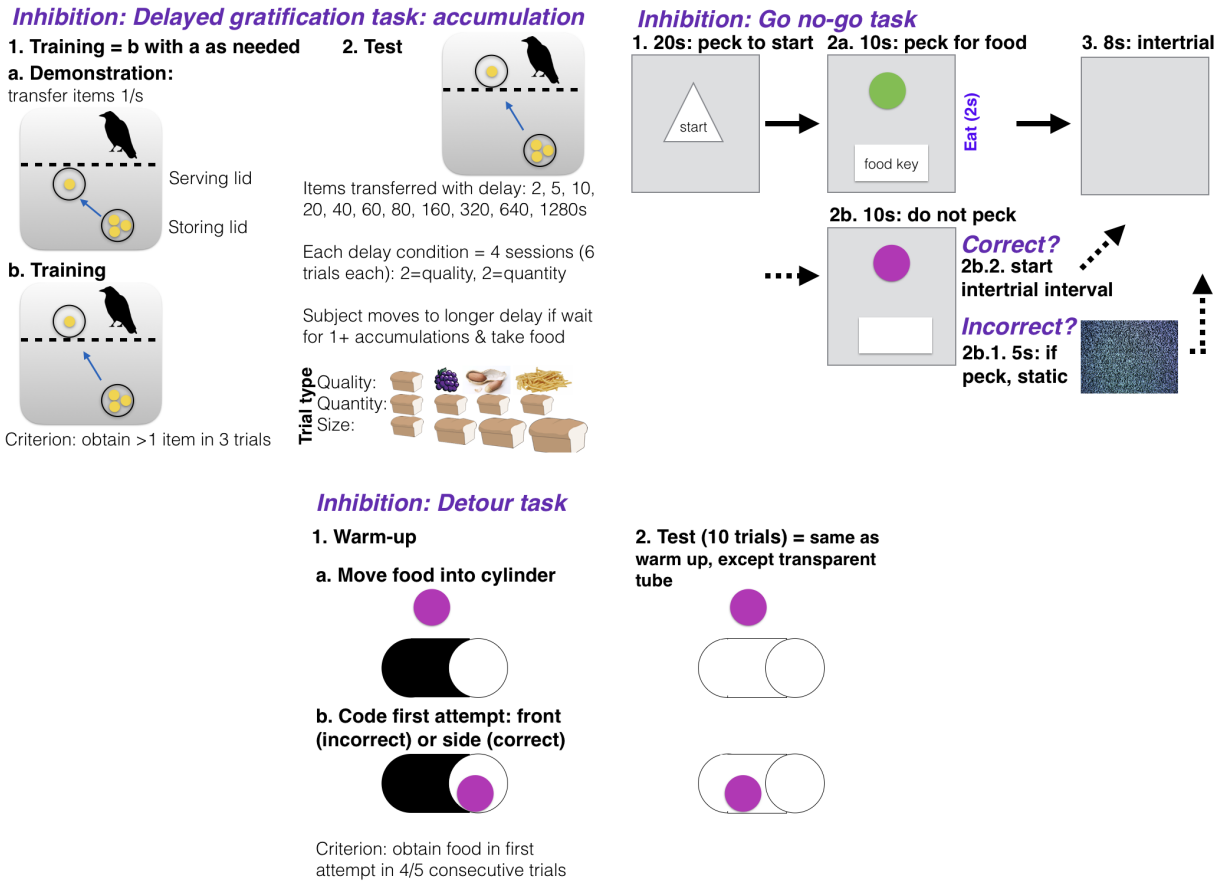


Figure 2: Figure 2. Experimental design.

goes blank for 8 seconds before starting over again. If the non-rewarded colored circle (purple in the figure) appears on the screen after the start key is pecked, then the correct response is to refrain from pecking the food key for 10 seconds. If the bird succeeds in refraining, the next intertrial interval starts. If the bird fails and pecks the food key while the purple circle is on the screen, then it is given an aversive stimuli for 5 seconds (TV static screen). In the **detour** task, individuals first receive a warm up with an opaque tube where they learn that the experimenter will show them a piece of food and then move that piece of food into the tube. They then have the opportunity to approach the tube and eat the food. A correct response is when their first approach is to go to the side of the tube to the opening to obtain the food and an incorrect response is when they try to access the food by pecking at the front of the tube (which has no opening). Once they pass the warm up, they move on to the test, which is exactly the same except the tube is transparent. The idea is that being able to see the food through the tube wall might entice them to try to go through the wall rather than refrain from a direct approach to the food and instead go around the side through the tube opening.

D. METHODS

Open materials

Testing protocols for all three experiments: color tube reversal learning, multi-access box, and touch screen reversal learning

Open data

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

Randomization and counterbalancing

P3

Two individuals from each batch will experience the detour task before participating in the flexibility manipulation. These individuals will be randomly selected using the random number generator at https://www.random.org.

P1-P2

For the rest of the individuals (n=6 per batch), the order of the three inhibition tasks will be counterbalanced across birds (using https://www.random.org to randomly assign individuals to one of three experimental orders). 1/3 of the individuals will experience:

1. Delayed gratification task
2. Go no-go task
3. Detour

1/3 of the individuals will experience:

1. Go no-go task
2. Detour
3. Delayed gratification task

1/3 of the individuals will experience:

1. Detour
2. Delayed gratification task

3. Go no-go task

Delayed gratification

- Food preference test: food will be presented in random combinations over six sessions of 12-15 trials.
- Training trials: The type of demonstration and training trials varied randomly (with more demo trials near the beginning of training), incorporating trials in which food of the same sort accumulated (quantity), food of ascending quality accumulated (quality), and trials in which we added increasingly larger food pieces throughout the trial (size)
- Test: we will test each food quality (low, mid, high) twice in randomized order in each session.

Go no-go

Go and no-go trials will be presented randomly with the restriction that no more than four of the same type will occur in a row. The rewarded color will be counterbalanced across birds.

Detour

The side from which the apparatus is baited will be consistent within subjects, but counterbalanced across subjects.

Blinding of conditions during analysis

No blinding is involved in this study.

Dependent variables

P1: the more flexible individuals are better at inhibition

1) **Delayed gratification:** Number of food pieces waited for (0-3). A successful wait is defined as waiting for at least one additional piece of food to be added to the serving lid of the three possible additional food items, and accepted at least one piece of the reward pieces.
2) **Go no-go:**

a) The number of trials to reach criterion (85% correct) where correct responses involve pecking when the rewarded stimulus is displayed and not pecking when the unrewarded stimulus is displayed, and incorrect responses involve pecking when the unrewarded stimulus is displayed, and not pecking when the rewarded stimulus is displayed

b) The latency to respond (peck the target key)

3) **Detour:** First approach (physical contact): Correct (to the tube's side opening) or Incorrect (to the front unopen area of the tube) (methods as in MacLean et al. (2014)).

One model will be run per dependent variable.

P3: does training improve detour performance?

1) First approach (physical contact): Correct (to the tube's side opening) or Incorrect (to the front unopen area of the tube) (methods as in MacLean et al. (2014)).

Independent variables

P1: delayed gratification

1) Food quality or quantity (Quality: High, Med, Low; Quantity: Smaller, Medium, Larger)
2) Trial
3) Delay (2, 5, 10, 20, 40, 60, or 80 seconds)
4) Flexibility 1: **Number of trials to reverse** a preference in the last reversal an individual experienced (reversal learning; 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.
5) Flexibility 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.
6) Flexibility 3: If the number of trials to reverse a preference does not positively correlate with the latency to attempt or solve new loci on the multi-access box (an additional measure of behavioral flexibility), then the **average latency to solve** and the **average latency to attempt** a new option on the multi-access box will be additional dependent variables. See behavioral flexibility preregistration.
7) Flexibility 4: 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 flexibility measures 1 through 3. If this ends up being the case, we will modify the code in the analysis plan below to reflect this change.

P1: go no-go

Model 2a: number of trials to reach criterion

1) Flexibility 1: Number of trials to reverse a preference in the last reversal an individual experienced (reversal learning; as above)
2) Flexibility 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 (reversal learning; as above).
3) Flexibility 3: If the number of trials to reverse a preference does not positively correlate with the latency to attempt or solve new loci on the multi-access box, then the **average latency to solve** and the **average latency to attempt** a new option on the multi-access box will be additional independent variables (as above).
4) Flexibility 4: 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 flexibility measures 1 through 3. If this ends up being the case, we will modify the code in the analysis plan below to reflect this change.

Model 2b: latency to respond

1) Correct or incorrect response

2) Trial

3) Flexibility Condition: control, flexibility manipulation

4) ID (random effect because multiple measures per bird)

P1: detour

1) Trial

2) Flexibility 1: Number of trials to reverse a preference in the last reversal an individual experienced (reversal learning; as above)

3) Flexibility 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 (reversal learning; as above).

4) Flexibility 3: If the number of trials to reverse a preference does not positively correlate with the latency to attempt or solve new loci on the multi-access box, then the **average latency to solve** and the **average latency to attempt** a new option on the multi-access box will be additional independent variables (as above).

5) Flexibility 4: 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 flexibility measures 1 through 3. If this ends up being the case, we will modify the code in the analysis plan below to reflect this change.

P3: does training improve detour performance?

1) Condition: pre- or post-reversal learning tests

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 3.3.3; 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.

Data checking

The data will be visually checked to determine whether they are normally distributed via two methods: 1) normality is indicated when the histograms of actual data match those with simulated data (Figure 2), and 2) normality is indicated when the residuals closely fit the dotted line in the Normal Q-Q plot (Figure 3) (Zuur, Ieno, and Saveliev 2009).

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

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

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


# Check the dependent variables for normality: Histograms
op <- par(mfrow = c(2, 3), mar = c(4, 4, 2, 0.2))
# This is what the distribution of actual data looks like
hist(acc$NumberOfAccumulationsWaited, xlab = "Delay: Number of accumulations waited",
    main = "Actual Data")
hist(go$TrialsToCriterion, xlab = "Go no-go: Trials to criterion",
    main = "Actual Data")
hist(detour$FirstApproach, xlab = "Detour: First approach", main = "Actual Data")

# Given the actual data, this is what a normal distribution
# would look like
X2 <- rnorm(1281, mean = mean(acc$NumberOfAccumulationsWaited),
    sd = sd(acc$NumberOfAccumulationsWaited))
hist(X2, xlab = "Delay: No. accumulations waited", main = "Simulated Data")

Y2 <- rnorm(1281, mean = mean(go$TrialsToCriterion), sd = sd(go$TrialsToCriterion))
hist(Y2, xlab = "Go/no-go: Trials to criterion", main = "Simulated Data")

Z2 <- rnorm(1281, mean = mean(detour$FirstApproach), sd = sd(detour$FirstApproach))
hist(Z2, xlab = "Detour: First approach", main = "Simulated Data")


# Check the dependent variables for normality: Q-Q plots
op <- par(mfrow = c(3, 4), mar = c(4, 4, 2, 0.2))
plot(glm(acc$NumberOfAccumulationsWaited ~ acc$Delay))
plot(glm(go$TrialsToCriterion ~ go$TrialsToReverseLast))
plot(glm(detour$FirstApproach ~ detour$Trial))
```

If the data do not appear normally distributed, visually check the residuals. If they are patternless, then assume a normal distribution (Figure 4) (Zuur, Ieno, and Saveliev 2009).

```
# Check the dependent variables for normality: Residuals
detour <- read.csv("/Users/corina/GTGR/data/data_detour.csv",
    header = T, sep = ",", stringsAsFactors = F)

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

go <- read.csv("/Users/corina/GTGR/data/data_go.csv", header = T,
```

```
    sep = ",", stringsAsFactors = F)

# Figure 3. Visual check of the residuals
op <- par(mfrow = c(1, 3), mar = c(4, 4, 2, 0.2))
plot(residuals(glm(detour$FirstApproach ~ detour$Trial)), ylab = "Detour residuals: First approach ~ Tri
plot(residuals(glm(acc$NumberOfAccumulationsWaited ~ acc$Delay)),
    ylab = "Delay residuals: Number of accumulations waited ~ Delay")
plot(residuals(glm(go$TrialsToCriterion ~ go$TrialsToReverseLast)),
    ylab = "Go/no-go: Residuals Correct response ~ Trial")
```

P1: delayed gratification

Assess food preferences: Conduct preference tests between pairs of different foods. Rank food preferences into three categories (High, Medium, Low) in the order of the percentage of times a food was chosen.

Analysis: Generalized Linear Model (GLM; glm function, stats package) with a Poisson distribution and log link, unless the only choices made were 0 (they didn't wait for food) and 1 (they waited for 1 piece of food but not for 2 or 3), in which case we will use a binomial distribution with a logit link. We will determine whether an independent variable had an effect or not using the Estimate in the full model.

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 $f^2 = 0{,}41$

err prob = 0,05

Power (1- err prob) = 0,7

Number of predictors = 5

Output:

Noncentrality parameter = 13,1200000

Critical F = 2,5867901

Numerator df = 5

Denominator df = 26

Total sample size = 32

Actual power = 0,7103096

This means that, with our sample size of 32, we have a 71% chance of detecting a large effect (approximated at f^2=0.35 by Cohen (1988)).

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

# GLM
better <- glm(NumberOfAccumulationsWaited ~ Delay + FoodQualityQuantity +
    Trial + TrialsToReverseLast + FlexRatio, family = "poisson",
    data = acc)
# summary(better)
```

```
better1 <- summary(better)
library(xtable)
better1.table <- xtable(better1)
library(knitr)
kable(better1.table, caption = "Table U: Model selection output.",
    format = "html", digits = 2)
```

P1: go no-go

Analysis:

Model 2a: Generalized Linear Model (GLM; glm function, stats package) with a Poisson distribution and a log link. We will determine whether an independent variable had an effect or not using the Estimate in the full model.

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

err prob = 0,05

Power (1- err prob) = 0,7

Number of predictors = 2

Output:

Noncentrality parameter = 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)).

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

# GLM
go1 <- glm(TrialsToCriterion ~ TrialsToReverseLast + FlexRatio,
    family = "poisson", data = go)
sgo1 <- summary(go1)
library(xtable)
sgo1.table <- xtable(sgo1)
library(knitr)
kable(sgo1.table, caption = "Table T: Model selection output.",
    format = "html", digits = 2)
```

Model 2b: A Generalized Linear Mixed Model (GLMM; MCMCglmm function, MCMCglmm package; (J. D. Hadfield 2010)) will be used 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). I will ensure the GLMM shows acceptable convergence (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 = 0,32

err prob = 0,05

Power (1- err prob) = 0,7

Number of predictors = 3

Output:

Noncentrality parameter = 10,2400000

Critical F = 2,9466853

Numerator df = 3

Denominator df = 28

Total sample size = 32

Actual power = 0,7061592

This means that, with our sample size of 32, we have a 71% chance of detecting a large effect (approximated at f^2=0.35 by Cohen (1988)).

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

# GLM
library(MCMCglmm)
prior = list(R = list(R1 = list(V = 1, nu = 0), R2 = list(V = 1,
    nu = 0), R3 = list(V = 1, nu = 0)), G = list(G1 = list(V = 1,
    nu = 0)))

golat <- MCMCglmm(LatencyToRespond ~ CorrectResponse * Trial *
    FlexibilityCondition, random = ~ID, family = "poisson", data = go,
    verbose = F, prior = prior, nitt = 13000, thin = 10, burnin = 3000)
summary(golat)
autocorr(golat$Sol)  #Did fixed effects converge?
autocorr(golat$VCV)  #Did random effects converge?
```

P1: detour

Analysis: Generalized Linear Model (GLM; glm function, stats package) with a binomial distribution and a logit link. We will determine whether an independent variable had an effect or not using the Estimate in the full model.

See the protocol for the power analyses for Model 2b above for the rough estimation our ability to detect actual effects with this model.

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

# GLM
detour$ID <- factor(detour$ID)
de <- glm(FirstApproach ~ Trial + TrialsToReverseLast + FlexRatio,
    family = "binomial", data = detour)
sde <- summary(de)
library(xtable)
sde.table <- xtable(sde)
library(knitr)
kable(sde.table, caption = "Table T: Model selection output.",
    format = "html", digits = 2)
```

P1 alternative 2: are inhibition results reliable?

The reliability of the inhibition tests will be calculated using Cronbach's Alpha (as in Friedman and Miyake (2004); R package: psych (Revelle 2017), function: alpha), which is indicated by std.alpha in the output.

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

library(psych)
reliab <- alpha(rel, check.keys = TRUE)  #Check.keys automatically reverses the coding for variables th
summary(reliab)
# Insert into text: `r reliab$std.alpha`
```

When comparing all three tests, alpha= .

P2: correlation across inhibition tasks

See analysis description for P1 alternative 2.

```
rel2 <- read.csv("/Users/corina/GTGR/data/data_inhibition2.csv",
    header = T, sep = ",", stringsAsFactors = F)

library(psych)
reliab2 <- alpha(rel2, check.keys = TRUE)
summary(reliab2)
# Insert into text: `r reliab2$std.alpha`
```

When analyzing only the delayed gratification and go no-go tasks, the reliability is alpha= *fill in result when data are available.*

P3: does training improve detour performance?

Analysis: Generalized Linear Model (GLM; glm function, stats package) with a binomial distribution and a logit link. We will determine whether an independent variable had an effect or not using the Estimate in the full model.

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 = 0,21

err prob = 0,05

Power (1- err prob) = 0,7

Number of predictors = 1

Output:

Noncentrality parameter = 6,7200000

Critical F = 4,1708768

Numerator df = 1

Denominator df = 30

Total sample size = 32

Actual power = 0,7083763

This means that, with our sample size of 32, we have a 71% chance of detecting a medium effect (approximated at f^2=0.15 by Cohen (1988)).

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

# GLM
de <- glm(FirstApproach ~ Condition, family = "binomial", data = detour)
sde <- summary(de)
library(xtable)
sde.table <- xtable(sde)
library(knitr)
kable(sde.table, caption = "Table T: Model selection output.",
    format = "html", digits = 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. 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).

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 will be 16, however we expect to be able to test up to 32 grackles.

Data collection stopping rule

We will stop testing birds once we have completed two full aviary seasons (likely in March 2020).

G. 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)

H. AUTHOR CONTRIBUTIONS

Logan: Hypothesis development, experimental design (go no-go task), data collection, data analysis and interpretation, write up, revising/editing, materials/funding.

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

Johnson-Ulrich: Touchscreen programming for go no-go task, data interpretation, revising/editing.

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

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

Seitz: Experimental design (go no-go task), touchscreen programming (go no-go task), data interpretation, revising/editing.

Blaisdell: Experimental design (go no-go task), data interpretation, revising/editing.

Wascher: Hypothesis development, experimental design (delayed gratification and detour tasks), data analysis and interpretation, write up, revising/editing.

I. 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.

J. ACKNOWLEDGEMENTS

We thank Dieter Lukas for help polishing the predictions; 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; and Richard McElreath for project support.

J. REFERENCES

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. Elsevier: 273–83.

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). Springer: 1149–60.

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). Springer: 175–91.

Friedman, Naomi P, and Akira Miyake. 2004. "The Relations Among Inhibition and Interference Control Functions: A Latent-Variable Analysis." *Journal of Experimental Psychology: General* 133 (1). American Psychological Association: 101.

Griffin, Andrea S, and David Guez. 2014. "Innovation and Problem Solving: A Review of Common Mechanisms." *Behavioural Processes* 109. Elsevier: 121–34.

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/.

Hadfield, JD. 2014. "MCMCglmm Course Notes." http://cran.r-project.org/web/packages/MCMCglmm/vignettes/CourseNotes.pdf.

Harding, Emma J, Elizabeth S Paul, and Michael Mendl. 2004. "Animal Behaviour: Cognitive Bias and Affective State." *Nature* 427 (6972). Nature Publishing Group: 312–12.

Hillemann, Friederike, Thomas Bugnyar, Kurt Kotrschal, and Claudia AF Wascher. 2014. "Waiting for Better, Not for More: Corvids Respond to Quality in Two Delay Maintenance Tasks." *Animal Behaviour* 90. Elsevier: 1–10.

Lefebvre, Louis, Patrick Whittle, Evan Lascaris, and Adam Finkelstein. 1997. "Feeding Innovations and Forebrain Size in Birds." *Animal Behaviour* 53 (3). Elsevier: 549–60.

MacLean, Evan L, Brian Hare, Charles L Nunn, Elsa Addessi, Federica Amici, Rindy C Anderson, Filippo Aureli, et al. 2014. "The Evolution of Self-Control." *Proceedings of the National Academy of Sciences* 111 (20). National Acad Sciences: E2140–E2148.

McElreath, Richard. 2016. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. CRC Press. http://xcelab.net/rm/statistical-rethinking/.

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. doi:10.1098/rsfs.2016.0121.

Peer, Brian D. 2011. "Invasion of the Emperor's Grackle." *Ardeola* 58 (2). BioOne: 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.

Revelle, William. 2017. *Psych: Procedures for Psychological, Psychometric, and Personality Research*. Evanston, Illinois: Northwestern University. https://CRAN.R-project.org/package=psych.

Sol, Daniel, and Louis Lefebvre. 2000. "Behavioural Flexibility Predicts Invasion Success in Birds Introduced to New Zealand." *Oikos* 90 (3). Wiley Online Library: 599–605.

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). National Acad Sciences: 5460–5.

Sol, Daniel, Sarah Timmermans, and Louis Lefebvre. 2002. "Behavioural Flexibility and Invasion Success in Birds." *Animal Behaviour* 63 (3). Elsevier: 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). Wiley Online Library: 1593–1607.

Zuur, Alain F.., Elena N.. Ieno, and Anatoly A Saveliev. 2009. *Mixed Effects Models and Extensions in Ecology with R*. Springer.

PROTOCOLS for Experiments: Delayed gratification, Detour, and Go No-Go

Dr. Corina Logan, University of Cambridge, cl417@cam.ac.uk, www.CorinaLogan.com
Dr. Claudia Wascher, Anglia Ruskin University

For hypotheses, predictions, methods, and analyses, see preregistration at:
https://github.com/corinalogan/grackles/blob/master/g_inhibition.Rmd

Counterbalancing across tests

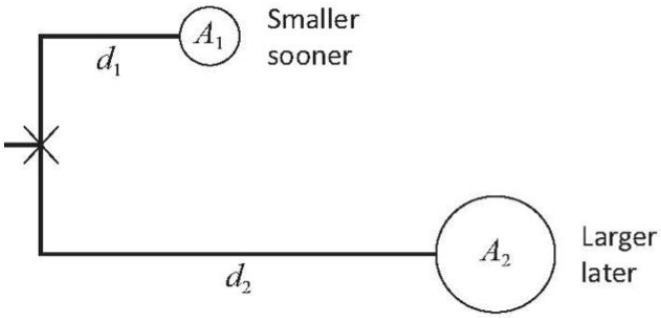
Table 1. Two individuals from each batch will experience the detour task before participating in the flexibility manipulation. These two individuals were randomly selected from each batch using random.org (n=8). The rest of the individuals (n=6 per batch) will be randomly assigned to one of three combinations of the order of inhibition experiments, counterbalancing the order such that ⅓ of the individuals experience Combination One: 1) delayed gratification, 2) go/no-go, and then 3) detour; ⅓ of the individuals experience Combination Two: 2, 3, 1; and ⅓ of the individuals experience Combination Three: 3, 1, 2 (random numbers generated at www.random.org).

Batch	Bird	Experimental order as in Combination	Batch	Bird	Experimental order as in Combination
1	1	3	3	1	3
1	2	1	3	2	2
1	3	2	3	3	2
1	4	3	3	4	Detour pre-reversal
1	5	Detour pre-reversal	3	5	3
1	6	Detour pre-reversal	3	6	Detour pre-reversal
1	7	2	3	7	1
1	8	1	3	8	1
2	1	3	4	1	3
2	2	3	4	2	1
2	3	2	4	3	2
2	4	2	4	4	Detour pre-reversal
2	5	Detour pre-reversal	4	5	1
2	6	Detour pre-reversal	4	6	Detour pre-reversal
2	7	1	4	7	2
2	8	1	4	8	3

1) Delay of gratification / Impulse control (after Hillemann et al. 2014)

Background:

Inter-temporal choices (Figure 1):



adapted from Stevens & Stephens 2009

Figure 1. Adapted from Stephens & Stephens 2009.

Self-control: ability to resist prompt gratification for a later benefit

Impulsive behaviour: choosing the sooner, but less valuable, reward

Delay of gratification: postponing an immediate response to a present reward to gain a more valuable reward in future

Delay choice: the initial selection between a delayed or an instantaneous reward; tasks: lever/button pressing, rotating tray (Addessi et al 2013)

Delay maintenance: maintaining one's decision to delay gratification, even if the instantaneous reward is made available later during the delay; tasks: accumulation, exchange (e.g. Hillemann et al 2014, Dufour et al 2012, Auersperg et al 2013)

Methods - The accumulation task:

1. *Food preference test:*

Determine individual food preferences. Two equally sized pieces of different types of food are presented simultaneously. The item to which the bird points the beak is given to the bird and rated as preferred food item, the not preferred food is removed. This is crucial as at the start of the experiments might not 'know' that the have to choose and will only get one piece of food. In the first couple of trials it often seems as if the bird would not actually take the preferred option but ideally the bird quickly learns that it only will get one piece and then really chooses the preferred option.

Use six types of food: bread, lettuce, grapes, maintenance diet, goldfish, Ritz (6 food types = 15 pairwise combinations, Table 2). They were tested in each possible combination three times (45 trials in total), randomly spread within 3 sessions of 15 trials. Each bird experiences trial sequence 1, then 2, then 3. For each individual a discrete hierarchy of three different qualities was determined: Low, Med, High.

For the preference test we did not have any criterion and just calculated the % of trials each food type was chosen.

Table 2. Order of the food preference trials. Numbers for food pairs: 1=lettuce, 2=bread, 3=grapes, 4=maintenance diet, 5=goldfish, 6=ritz. Food pair combinations were randomly dispersed in three 15-trial sequences such that each possible food pair combination was experienced three times (random numbers generated at www.random.org). Each bird runs through all

three trial sequences and thus has exposure to each possible food combination three different times. This allows for a more comprehensive understanding as to which food the birds prefer in each combination.

Combination number	Food pair	Trial sequence 1 (combination number)	Trial sequence 2 (combination number)	Trial sequence 3 (combination number)
1	1-2	6	4	7
2	1-3	1	5	8
3	1-4	12	3	15
4	1-5	5	7	5
5	1-6	11	6	14
6	2-3	7	2	12
7	2-4	4	15	9
8	2-5	10	11	10
9	2-6	8	8	2
10	3-4	9	12	11
11	3-5	15	9	1
12	3-6	3	1	13
13	4-5	2	10	6
14	4-6	13	14	3
15	5-6	14	13	4

1. *Training ('behaviour shaping'):*

The set-up consisted of two plastic lids (3.8 cm diameter [we used lids of probiotics supplements], one storing ('out of reach container') and one offering ('in-reach container') the reward.

Wire-mesh: we used the wire-mesh in the front of the aviary as a barrier between the experimenter and the bird, with the storing cup on the experimenter's side of the mesh and the serving cup on the bird's side so the bird could reach the food.

At the start of a trial, the experimenter shows the content of the storing lid to the subject and set it next to the serving lid so that all food items remained visible but out of reach for the bird. Per trial, the subject could gain up to four pieces of food, either of the same food type or of increasing quality.

-Demonstration trials: To demonstrate the procedure, the experimenter placed both plastic lids out of the bird's reach while transferring the items from the storing into the serving lid without delay, that is, the reward grew at a rate of one item every second. When all four pieces had been transferred, the experimenter moved the serving lid within reach of the bird, so it could take the food.

-Training trials: In contrast to demonstrations, in training trials the experimenter placed the serving lid in reach of the subject, so it was able to take the food at any time during a trial. As in the demonstrations, the experimenter delivered food items without delay. As soon as the subject touched an item, the accumulation ended by the experimenter removing the storing lid. During the inter-accumulation interval (a few seconds), the bird was able to eat or take the food. If the subject waited until the accumulation ended, this meant that all four pieces of food were in place and able to be reached by the subject; the reward then remained there for 5 s to give the bird the chance to take it. We noted the items that the subject did not take as being refused (if a bird does not take the reward this is an issue for the experiment (lack of motivation), because it is easy to

wait if you do not want it in the first place. Ideally it never happens that the bird does not take the reward, if it happens regularly, experimental design needs to be adapted (e.g. less regular testing, changes in food type).

The training phase consisted of 8-12 sessions with approximately 5-10 trials, demonstrating and training the method. The type of the demonstrations and training trials varied randomly, incorporating trials in which food of the same sort accumulated (quantity), food of ascending quality accumulated (quality), and trials in which we added increasingly larger food pieces throughout the trial (size). As assumed by Anderson et al. (2010), and Evans and Beran (2007), the gradually increasing attractiveness of ascending sizes helps individuals to understand the task intuitively.

Initial sessions consisted of demonstration trials, whereas in later training sessions we increased the number of training trials while we reduced the number of demonstrations to three per session: the first trial, a trial in the middle of the session, and the final trial.

Training criterion: subject succeeded in gaining more than one food item in at least three training trials.

Test: In the first condition, the food accumulated with a delay of 2 s between each piece. Delays consecutively increase over conditions: 2 s, then 5, 10, 20, 40, 60, 80, 160, 320, 640, 1280 s. If the subject succeeded in waiting at least once at a given stage, and did not refuse the reward, it proceeded to the subsequent stage; with the exception that we tested all subjects in the first two delay conditions, regardless of their performance in the 2 s condition. Each delay condition contained four sessions of 6 trials: two sessions of quality trials (sessions 1-2), and two sessions of quantity trials (sessions 3-4). In the quantity sessions (3 and 4), we tested each food quality (low, med, high) twice in randomized order (note: in quantity sessions, the food type within a trial stays the same, but different food qualities are tested in different trials). In the quantity sessions (3 and 4), each trial involves three pieces of the same type of food accumulating, whereas in the quality sessions (1 and 2), each trial involves three pieces of different quality foods accumulate from lowest quality to highest quality (first L, then M, then H). The quantity sessions (3 and 4) included two trials of each food type.

Table 1. Assigning random numbers to trial type and session for each bird.
Trial type: L=low quality food, M=medium quality, H=high quality. Session = 6 trials (2 of each trial type).
Randomized order of 3 trial types (L, M, H) within each session:
Session 3 (quantity): H, M, H, L, L, M.
Session 4 (quantity): H, M, H, L, M, L.

Batch	Bird	Order of sessions (same order for each delay)	Batch	Bird	Order of sessions (same order for each delay)
1	1	2, 4, 3, 1	3	1	2, 4, 1, 3
1	2	4, 3, 1, 2	3	2	2, 4, 1, 3
1	3	4, 1, 2, 3	3	3	2, 3, 1, 4
1	4	1, 2, 4, 3	3	4	1, 2, 4, 3
1	5	2, 4, 1, 3	3	5	3, 1, 4, 2
1	6	4, 1, 2, 3	3	6	2, 1, 4, 3
1	7	3, 2, 4, 1	3	7	3, 4, 1, 2
1	8	4, 1, 2, 3	3	8	4, 1, 3, 2
2	1	1, 3, 2, 4			
2	2	4, 1, 2, 3			
2	3	3, 1, 2, 4			
2	4	4, 3, 1, 2			
2	5	2, 4, 3, 1			
2	6	2, 1, 3, 4			

2	7	2, 3, 1, 4			
2	8	4, 3, 2, 1			

2) Detour task (experimental design after MacLean et al. 2014)

Food items were placed in a plastic cylinder (length = 15cm, diameter = 10.5cm) made from a 2-liter soda bottle. During the 'warm-up', the cylinder was opaque and during test trials it was transparent. The cylinder was mounted to a plastic base and set on the ground such that it remained stationary throughout the trial and so that subjects must adjust their approach toward the apparatus, rather than manipulate the position of the apparatus itself when retrieving food.

1. Warm-up:

In the warm up condition, a food reward is shown to the bird by holding it up above the apparatus. As the subject was looking at the experimenter, the food reward was placed inside the opaque, plastic cylinder. The side from which the apparatus was baited (from the point of view of the experimenter) was consistent within subjects but counterbalanced across subjects.

Once the food reward was positioned, the subject was allowed to approach and retrieve this item. If the subject did not approach within 30 seconds the cylinder was re-baited (i.e., the bait that was previously placed in the cylinder was removed, shown to the bird again, and then placed in the cylinder again). A trial was counted when a bird made a choice and the number of rebaitings was counted as an assessment of the number of potential learning opportunities. On every trial the experimenter coded whether the subject's first attempt to retrieve the item was by touching the front of the apparatus (incorrect) or entering from the side (correct: successful detour). Subjects were permitted to retrieve the food reward on all trials regardless of the accuracy of their first attempt.

Criterion: Subjects were required to correctly retrieve the food reward (on the first attempt) in 4 of 5 consecutive trials before receiving the test. Once this criterion was met, subjects advanced directly to test trials. The warm-up was limited to 10 trials, however if a subject needed more trials to meet criterion, additional trials were given.

2. Test:

The test procedure was identical to warm-up trials except that the apparatus used was the transparent cylinder. 10 trials were conducted with all subjects regardless of performance during test. As in warm-up trials, the experimenter coded whether the subject's first attempt to retrieve the item was through the front (incorrect) or side (correct) of the apparatus. Again, subjects were allowed to retrieve the item on all trials regardless of the accuracy of their first attempt.

Table 2. Randomly assigning birds to experience observing the food placed into the tube from the left (1) or right (2), from the perspective of the experimenter, such that half of the birds from each batch experience one side.

Batch	Bird	Side the food is observed to enter	Batch	Bird	Side
1	1	Left	3	1	Right
1	2	Right	3	2	Left
1	3	Right	3	3	Right
1	4	Right	3	4	Right
1	5	Left	3	5	Right
1	6	Left	3	6	Left
1	7	Right	3	7	Left
1	8	Left	3	8	Left
2	1	Left	4	1	Right

2	2	Right	4	2	Right
2	3	Left	4	3	Left
2	4	Left	4	4	Left
2	5	Left	4	5	Left
2	6	Right	4	6	Left
2	7	Right	4	7	Right
2	8	Right	4	8	Right

3) Go/no-go task (experimental design after Harding et al. 2004)

The apparatus: a touch screen with an automated food hopper below it containing peanut pieces. On the screen is a *trial start key* (white triangle, 4cm sides), a *food key* (white square, 4cm by 4cm tall) that they can peck to respond (or not) to the stimulus displayed above it. The *stimulus* is a circle (4cm diameter) in different colors.

Training: training for most individuals will occur before reversal learning on the touch screen and therefore they will not undergo training again, but move directly on to the test. Those individuals in the flexibility control group (see flexibility preregistration for details: https://github.com/corinalogan/grackles/blob/master/g_flexmanip.Rmd), will now receive the same training as individuals in the manipulation group on how to use the touch screen (described in Experiment 3 at https://docs.google.com/document/d/18D80XZV_XCG9urVzR9WzbfOKFprDV62v3P74upu01xU/edit?usp=sharing and listed here).

Training: food hopper - Generally, slowly decrease the time the hopper is available and the speed with which it is withdrawn and returned. We would like grackles to associate the sound of the hopper moving with food being available (note: a light also turns on when the food hopper is available, however this experiment is conducted in outdoor aviaries where it is bright and thus the light might not be the most obvious cue). End goal behavior for hopper training: grackle lands on platform, hopper is moved forward within reach, grackle retrieves food, hopper is moved out of reach, grackle is able to retrieve food immediately and passes criterion (point 4 below).

1. Position the food hopper so it is in the accessible position. Draw attention to it by placing peanut crumbs around the area. Allow the bird to eat from the hopper for 20 seconds, then go into aviary and add more crumbs at/around hopper. Repeat until the bird eats from the hopper without the peanut crumbs.
2. The experimenter uses the Maestro program to slowly pull the hopper out of reach for 1-5 seconds (so that the grackle has noticed food is not longer available), then slowly moves the hopper back into a reachable position. Allow grackle to eat for 5-20 seconds (or until it is seent with 3 food items in its bill, so it has eaten at least 3 food items). Repeat. Set it so the food stays available at first for 20s and gradually decrease it to 1-3 seconds (which is what it will be during testing). Gradually increase this speed until the grackle does not retreat or show signs of fear (e.g., flying away, jumping backwards, reluctant to return to hopper, reluctant to put head in hopper). If grackle leaves the platform, make the food unavailable and only return it when the grackle is on the platform facing the hopper.
3. ***Criterion:*** Continue until grackle is able to retrieve food immediately in 8 out of the most recent 10 opportunities (as in Bateson et al. 2015 PLOS ONE) with the hopper moving forward and backward at maximum speed. Intertrial interval = 5-10s (food is not available during this time so the grackle learns it must pay attention to when it is available).

Training: touch screen

- *Passing **criterion** for each training program:* Subject needs at least 17 of the most recent 20 trials correct (touch the screen and eat from the hopper after each correct touch) with at least 8/10 or 9/10 correct in the most recent 2 sessions (each consisting of 10 trials). After each session, check the data file for the number of correct trials.

1. ***Initiate interest in pecking touchscreen:*** Begin with **"moving dot**" program. The purpose of this training is to use an ecologically relevant stimulus (similar to a flying insect, which grackles regularly forage on) to facilitate faster learning of an association between pecking a stimulus on the screen (which also results in audio feedback) and obtaining a food reward. The correct response is to peck a small, moving white stimulus (diameter=0.5cm), which results in the receipt of a food reward. Once criterion is met, move the bird on to the "Position Shape" training program.

2. ***Peck white square for food:*** Conduct "**position shape**" program. The purpose of this training is for the bird to learn to peck a non-moving stimulus (a white square), that appears in different locations on the screen (between trials, rather than within trials) to trigger the food hopper. The correct response is to peck the white square and eat the food. Once criterion is met, move the bird on to the "Start Key Shape" training program.
3. ***Peck trial start key to initiate white square:*** Conduct "**start key shape**" program.The purpose of this training is for the bird to learn to peck the trial start key (white triangle) which, during testing, will not result in a food reward, but will result in the beginning of a trial. This program starts with a flashing white triangle (on for 0.7s, off for 0.3s as in Bateson & Kacelnik, 1995) centered at the bottom of the screen. Once pecked, the white triangle disappears and a white square appears in the triangle's location. The correct response is for the bird to peck the white square, which results in receipt of a food reward. Once criterion is met, begin test.

Test: color discrimination. Peck the rewarded color and don't peck the unrewarded color (color: pink or green circle counterbalanced across individuals)

1. Peck the start key (a *flashing key* on for 0.7s, off for 0.3s as in Bateson & Kacelnik 1995; above where the food key would be; white square 4x4cm) to initiate a trial (controls for motivation to participate in the task). It will flash for 20s and if the bird hasn't pecked it to start the trial, it will disappear and the screen will go blank for 1 min. Repeat.
2. Stimulus (green circle or pink circle) is present on screen for 10 seconds
    1. The bird will need to peck the food key when the *rewarded stimulus* is on screen. After the 10s, the screen goes blank and the bird waits for the next trial to start
    2. For the *unrewarded stimulus*, they will need to NOT peck the food key when the stimulus is present; if the bird pecks the screen outside the food key, nothing will happen (no food reward and no aversive stimulus). After these 10s, the screen goes blank and the bird waits for the next trial to start.
    3. If the bird pecks the food key when the *unrewarded stimulus* is present on the screen (during the 10s), then there is a penalty (TV static screen as an aversive stimulus for 5s).
3. 8 second intertrial intervals (mean +/-4s) start at the end of food delivery or at the end of the TV static screen.
4. Continue to test until the bird reaches 100% accuracy (or 85% accuracy if it takes more than 150 trials)

The rewarded stimulus and unrewarded stimulus are presented in randomized order but never with more than two of the same in a row. Each session consists of 20 trials and each bird can participate in up to four sessions per day. Sessions always start with a rewarded color trial (Bateson et al. 2015 PLOS ONE).