

# Do the more flexible individuals rely more on causal cognition? Observation versus intervention in causal inference in great-tailed grackles

Dr. Aaron Blaisdell (University of California Los Angeles), Dr. Zoe Johnson-Ulrich (University of California)

31 January 2019

```
#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)) (see an overview of the 5-year project timeline). **This investigation:** In this piece of the long-term project, we aim to determine whether the more behaviorally flexible (measured in a separate preregistration) grackles are better able to make causal inferences (understanding relationships beyond their statistical covariations) in two experiments using a touch screen apparatus. Results will indicate whether causal cognition might play a role in switching to functionally relevant solutions based on how it correlates with measures of flexibility (reversal learning and solution switching). This will improve our understanding of which variables are involved in 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, underwent two rounds of peer reviews and was recommended at PCI Ecology (see the review history).

## B. PARTITIONING THE RESULTS

We may decide to present these results in two separate papers: 1) determining whether grackles have causal inference abilities, and 2) linking variation in performance on causal inference tasks to measures of flexibility.

## C. HYPOTHESIS

#### Individuals that are more behaviorally flexible (faster at functionally changing their behavior when circumstances change), as measured by reversal learning and switching between options on a multi-access box, are better able to derive accurate causal inferences (see Mikhalevich, Powell, and Logan (2017) for theoretical background about the distinction between flexibility and complex cognition). This is because causal cognition may facilitate flexibility: an individual could be faster at switching to new solutions that are more functional if it makes causal inferences about how the problem works, rather than relying solely on trial and error learning to indiscriminately switch to new solutions. In this procedure, we assess whether grackles are able to derive correct predictions about causal interventions after observational learning, a core component of causal reasoning that can not be reduced to associative learning (Waldmann and Hagmayer (2005)).

**Predictions:** Individuals that are faster to reverse preferences on a serial reversal learning task and who also have lower latencies to successfully solve new tasks after previously solved tasks become unavailable on a multiaccess box (two measures of flexibility in a separate preregistration), perform better in two causal inference experiments. Specifically, the more flexible individuals are predicted to

- P1: form causal models from contingency learning (i.e. observational learning). Contingency information could be represented in one of two ways. On the one hand, relations between events could be encoded as associations. On the other, they could be represented as causal. For example, if the sound of a bell is followed by delivery of food, one could represent the bell as associated with the food, and thus the sounding of the bell calls to mind an expectancy of food. Or, the subject could represent the bell as a cause of food. Blaisdell et al. (Blaisdell et al. (2006)) (see also Leising et al. (2008)) report evidence that rats can represent statistical relationships between events as causal.” Thus, we predict the more flexible individuals will better learn the causal maps between all pairwise events (visual and auditory cues and food delivered from a food dispenser), and integrate these individual maps into larger causal map structures, including a common-cause, two-effect map (if observing T, L caused it, thus F is present), and a direct cause-effect map (if N is present, it will cause F).
- P2: behave as if intervention can influence the type of causal inference made at test, depending on which causal model is being tested: dissociate between seeing and doing as evidenced by a lower rate of pecking a key to release food when they had the opportunity to intervene in a common cause condition, while intervening on a direct cause or a causal chain will have no effect on key pecks.

Alternative 1: If there is no correlation between flexibility measures and performance on causal inference tasks, this suggests that learning about associations (on which the flexibility tasks are based) is different from learning about causal inferences.

Alternative 2: If there is a negative correlation between flexibility measures and performance on causal inference tasks, this suggests that some individuals may prefer to rely on information acquired previously (i.e., they are slow to reverse, but good at remembering prior information in causal inference tasks) rather than relying on current cues (e.g., the food is in a new location (flexibility), the light is absent in the test trials (causal inference)). For example, relying solely on current cues (i.e., the immediate stimulus (e.g., tone, noise) or lack thereof) in the causal cognition test will not give them enough information to consistently solve the task. They will need to draw on their memory of what the presence or absence of the current stimulus means about the food reward based on their experience in previous trials to perform well on this task.

Alternative 3: If the flexibility measures do not positively correlate with each other (P2 alternative 2 in the flexibility preregistration), this indicates they measure different traits. In this case, we are interested in how each flexibility measure relates to performance on causal inference tasks: the reversal learning measure as an indication of flexibility, and task switching latency on the multiaccess box as a measure of a combination of flexibility and innovation.

**Figure 1.** Test figures adapted from Blaisdell et al. (2006). In the Training phase, subjects first learn to peck at a food key to elicit food from the magazine. Subjects then receive trials during which the yellow circle is presented on the screen followed by a tone, and then they receive two types of trials interspersed

## Causal cognition

### Experiment 1: Form causal models from contingency learning: Common cause vs. Direct cause

A. Intervene Tone & B. Observe Noise (n=8)  
C. Intervene Noise & D. Observe Tone (n=8)

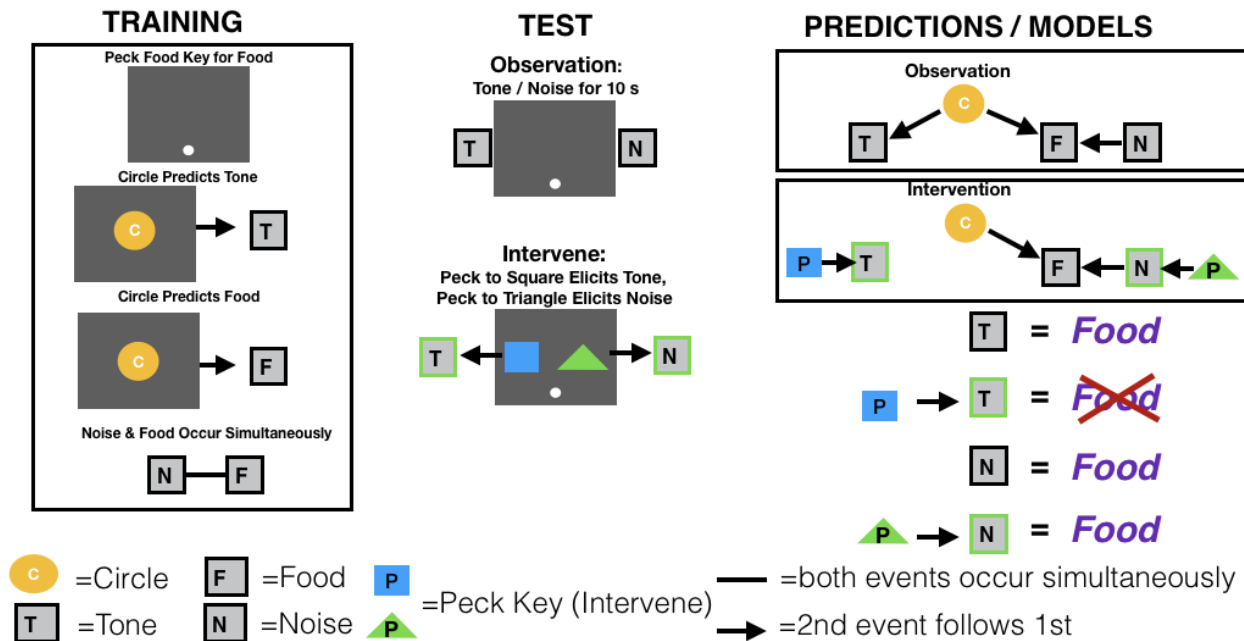


Figure 1: Figure 1. A visual illustration of the the training, test, and predictions in Experiment 1 (T=tone, L=light, F=food, N=noise, P=peck key). Test figure from Blaisdell et al. (2006).

within each training session: 1) the yellow circle followed by food or 2) the noise and food presented at the same time. In the Test phase, those individuals in the Observation condition will hear a tone or a noise while seeing only the food key on the screen but individuals in the intervene condition can elicit the tone or noise by pecking on separate response keys that elicit those auditory stimuli. The prediction is that if individuals form a common cause model such that circle produces tone and food, observing the tone should lead to individuals to expect food (indicated by the subject pecking the food key), because if tone is on, circle caused it, and it also caused food. Meanwhile, if they intervene to produce the tone, they will not expect food (i.e., they will not peck the food key) because they know their intervention caused the tone and not the circle (which also causes food). Additionally, observing noise should also lead an expectation of food (i.e., pecking the food key) because noise and food were paired simultaneously during training. Meanwhile, even if they intervene and cause the noise they should still expect food (i.e., peck the food key) because when the noise is on food is available regardless of what caused it (because the noise and the food were paired simultaneously).

## Causal cognition

### Experiment 2: Common cause (2a) vs. Causal chain (2b)

Individuals (different from those in Experiment 1) randomly assigned to:

- Intervene (2a n=4, 2b n=4)
- Observe (2a n=4, 2b n=4)

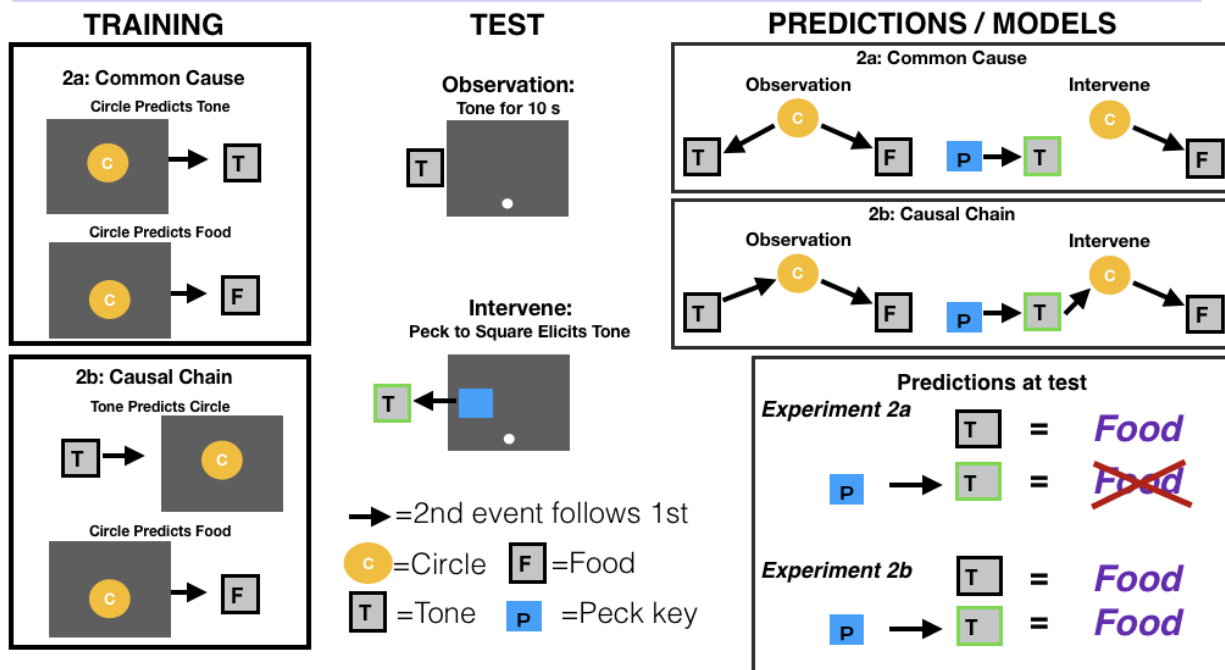


Figure 2: Figure 2. A visual illustration of the the training, test, and predictions in Experiment 2 (T=tone, L=light, F=food, P=peck key). Test figure from Blaisdell et al. (2006).

**Figure 2.** Test figures adapted from Blaisdell et al. (2006). In Training phase 2a, subjects learn the same common cause model as in experiment 1 such that the circle predicts tone and also the circle predicts food. Note that there is no training of the noise in 2a. In the training phase of 2b, subjects learn that a tone comes before a circle on screen and that a circle on screen is followed by receiving food. In the Test phase, those individuals in the Observation condition hear a tone while seeing only the food key on the screen and those in the intervene condition can elicit the tone by pecking at a blue square. For subjects trained in 2a, the prediction is that if individuals form a common cause model such that circle produces tone and food,

observing the tone should lead to individuals to expect food because, if tone is on, the circle caused it, and the circle also causes food. If individuals intervene to create the tone, they should not peck for food because it was their action that caused the tone and not the circle that caused it. For subjects trained in 2b, the prediction is that if they observe the tone they should expect food because food follows the circle which follows the tone. Likewise, even if subjects intervene to produce the tone, if they have formed the causal chain, they should still expect that the tone produces the circle which produces food and hence they should look for food.

### **Objective:**

The aim is to determine whether grackles, like rats (Blaisdell et al. (2006)), derive predictions about the outcomes of interventions (actions) after passive observational learning of different kinds of causal models. Causal models are theoretical entities that are estimated, combining cues to infer causal structure in relationships that go beyond merely observing statistical covariations between events.

Blaisdell and colleagues (Blaisdell et al. (2006)) taught rats that a light was a common cause of tone and food by presenting the light followed by the tone on some trials and by the food on other trials during training. Rats also learned that a noise was a direct cause of food by presenting noise and food simultaneously during training. At test, some rats observed the tone or the noise. When they did, they looked for food. This shows that rats had formed the causal models of noise causes food and that tone is caused by light, which itself is a cause of food. Other rats were given the opportunity to intervene to make the tone or noise occur at test. This was done by giving the rats a novel lever that they had never seen before or been trained on. When they pressed the lever, this caused the tone (or noise) to turn on. When the noise was caused by a lever press, rats looked for food in the food hopper, but when lever pressing caused the tone to turn on, rats did not look for food. This shows that rats understood that, by intervening on the lever to cause the noise to occur, since the noise was a cause of food, they then expected food. But by intervening on the lever to cause the tone to occur, the rats realized that they had caused the tone, and not the light (which was an alternative cause of tone). As a result of attributing the tone to their own action rather than the light, they did not expect there to be any food in the food hopper.

This experiment adapts the procedure used by Blaisdell et al. (2006) to study causal inference in rats for the study of causal inference in birds (e.g., pigeons and grackles) using a touchscreen. Blaisdell et al. (2006) (see also Leising et al. (2008)) found that rats made different predictions about the presence of food based on a cue (a tone) depending on the causal relationship between them (direct cause or two effects of a common cause) and whether the tone was merely observed at test or had been caused by the subject's own intervention (a lever press). This dissociation between seeing and doing suggests that subjects represent associated relationships as causal, and derive rational inference regarding an intervention on a cause versus an effect. We wish to determine whether grackles can also form causal models from contingency learning, and if so, whether their intervention can influence the type of causal inference made at test, depending on which causal model is being tested.

## **D. METHODS**

### **Planned Sample**

Great-tailed grackles will be caught in the wild in Tempe, Arizona USA for individual identification (colored leg bands in unique combinations). Some individuals (~32: ~16 per experiment) will be 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 three years we have at this field site given that the birds only participate in tests in aviaries only during the non-breeding season (approximately September - March). The minimum sample size will be 8 birds per experiment ( $n=16$  total), however we expect to be able to test many more.

## **Data collection stopping rule**

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

## **Open materials**

Touchscreen training protocol.

## **Apparatus**

Testing was conducted on an operant touchscreens mounted on a platform and placed on a cart inside an individual subject's aviary. All stimuli were presented by computer on a color LCD monitor (NEC MultiSync LCD1550M). Pecks to the monitor were detected by an infrared touchscreen (Carroll Touch, Elotouch Systems, Fremont, CA) mounted in front of the monitor. A food hopper (Coulbourn Instruments, Allentown, PA) was located below the monitor with an access hole situated flush with the floor. When in the raised position, the hopper provided access to peanut pieces. Laptop speakers delivered a pure tone or a clicking noise stimulus. Geometric symbols could be displayed on the touchscreen monitor. These consisted of a white star with gray lines (formerly yellow circle), and a square and triangle. The latter two shapes were both white with gray lines on them (formerly one was blue the other green) (the shape-sound pairings will be counterbalanced across subjects, while the star is for all subjects). All experimental procedures were programmed using PsychoPy (v1.85.2, Peirce (2009)).

## **Touch screen training**

For the most up to date touchscreen training protocol, please see our Google sheet.

## **Training: food hopper**

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). Every time the grackle approached the food hopper, we opened it remotely. The opening makes a distinct sound, and the food hopper is left open until the grackle looked in before closing. 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 can not obtain food.

1. Position the food hopper so it is in the accessible position. Draw attention to it by placing food 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 crumbs.
2. To habituate the bird to the sound of the hopper moving, use 1.Press\_Space\_for\_food\_2.Basic\_mag\_training\_.psyexp program which lets you press the spacebar to raise, and lower the hopper. Raise and lower the hopper many times when the bird is attending to the apparatus, especially when the bird is on the platform. Allow the bird to come to the platform and eat from the hopper, than immediately after press the spacebar to lower it while the bird is watching. Continue to do this until the bird is no longer jumpy.
3. To train the grackle to eat quickly from the hopper, the experimenter uses the 1.Press\_Space\_for\_food\_2.Basic\_mag\_training\_.psyexp program. Again use the spacebar to raise the hopper into the open position when the grackle is on the platform. Allow the food hopper to be available for eating for 20 s. After this time period, move the hopper out of reach. Wait 5-10 seconds (so that the grackle has noticed food is not longer available),

then initiate another trial to move the hopper back into a reachable position. Allow grackle to eat for 5-20 s (or until it is seen with 3 food items in its bill, so it has eaten at least 3 food items). Repeat. At first, let the food stay available for 20s and gradually decrease it to ~8 s (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. Once grackle has habituated to going to, and eating from, the food hopper when it opens, proceed with trials to assess whether grackle passes hopper training.

4. Open a new run on the PsychoPy hopper training program. Enter the bird's ID and S1 T1 to indicate the first session and first trial of hopper training trials (or subsequent sessions/trials as appropriate). Once program is running, turn on camera and use board to again indicate experiment, session, and trial number. Then put a small piece of cracker in front of the hopper before leaving the cage. Once the grackle comes to the platform, let it take the free piece of cracker, then press the spacebar to raise the hopper. Count to 5, if the grackle sticks its head in the hopper, let it eat up to 3 pieces of food. Lower the hopper and count to 10 before raising the hopper again for the next trial. In the data\_reversaw sheet, put a "1" in CorrectChoice if the grackle ate within 5 sec from the food hopper. If the grackle does NOT stick its head into the hopper within 5 seconds, lower the hopper. Count to 10 before raising the hopper again for the next trial. In the data\_reverseaw sheet, put a "0" in CorrectChoice if the grackle did NOT eat within 5 sec from the food hopper. If the grackle leaves the platform after you raise the hopper, put a "-1" in CorrectChoice and repeat the trial the next time the grackle comes to the platform. If the grackle left the platform and does not come back within 5 min of the start of the previous trial, end the session and try again after giving the grackle a break. Grackles should get 20 trials - ideally 2 sessions of 10 trials.
5. **Criterion:** Subject needs at least 17 of the most recent 20 trials correct (with the hopper moving forward and backward at maximum speed), with at least 8/10 or 9/10 correct in the most recent two 10 trial blocks (as in Bateson et al. (2015)). Intertrial interval = 10s (food is not available during this time so the grackle learns it must pay attention to when it is available). NOTE: if a bird passed criterion the previous day, re-run that program to make sure they retained the information before moving on to the next program.

## Training: touch screen

### General Notes:

- Once a bird is habituated to the food hopper and touchscreen, only put the touchscreen in their aviary when you want them to pay attention to it and, if they make the correct response, they get a reward. Because hand shaping works the best for training grackles on the touchscreen, all training sessions must be attended by the experimenter who must pay attention the entire time (see the video on "how Dazzle learns to blow bubbles" to learn about how hand shaping works: <http://www.dogtrainingology.com/concepts/shaping-behavior-definition/>).
- If a bird is having trouble with motivation/focus:
  - Take them back to the last program they were successful at and let them have a few good trials before trying the more advanced program. Also, if they are giving up on participating in a program because they are frustrated, end on a good note by taking them back to a program they already know and give them a few successful trials to make sure they stay interested in interacting with the touchscreen
  - Demonstrate by touching the screen how to use the program
  - If the bird touches the correct stimulus on the screen, but it doesn't activate due to a program error, use your cursor to click the button for them so they are rewarded for the correct behaviors and continue to progress with their learning.

**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.

## Peck food key for food

Program: 4.Food\_Key\_Only\_2FullControl.psyexp

How it works: a white square (3.5cm by 3.5cm) is on screen, when pecked, it disappears and the food hopper automatically becomes available. The experimenter must press the space bar to make the food unavailable, and then the white square re-appears. Experimenter can open or close the food hopper at any time by pressing space bar.)

Begin the program. To get the grackle interested in the screen, tape a goldfish cracker to the screen on top of the white square - Try to create a tape “hammock” so that the sides and bottom are taped on, but there is a gap in the center top where you can drop in crackers - Put a cracker piece in the tape “hammock”, when the grackle takes the piece press the spacebar to raise the hopper - Let the grackle eat 2-3 pieces from the hopper before pressing the spacebar to close it - If the grackle does not notice the hopper, close it after about 8 seconds. - If they don’t come back to the table, bait the tape again with a very small cracker crumb - Once the grackle is taking the cracker comfortably, remove the cracker from the screen - To get the grackle to associate the white square food key with food, hand shape by rewarding (giving them enough time to eat a couple of pieces of food) when the bird’s bill is near the white square, and then when it is closer to the square, and then when it is touching the square. The hopper automatically comes up when the bird pecks the square (but if it doesn’t and the bird gave the correct response, then press the space bar to move the hopper up), and then press the space bar to move the hopper out of reach after the bird gets one or two pieces of food.

## Experiment 1

Before contingency training, subjects completed response key autoshaping and instrumental conditioning. Subjects were trained to peck at the response key to activate the food hopper using a mixed autoshaping/instrumental training procedure.

Contingency training follows the procedure outlined in Figure 1. The light is a white star (4cm wide x 4.5cm tall), the tone is 400hz presented for 10s, and the noise is a clicking noise presented for 10s. In the first training phase, subjects will receive trials during which the star will be presented on the screen followed by presentation of the tone. The star and tone will each be presented for 10s with the onset of the tone coinciding with the termination of the star. During the second training phase, subjects will receive two types of trials interspersed within each training session. On some trials, the star will be presented for 10 seconds followed by the delivery of food from the hopper. On other trials, the 10s noise and 10s delivery of food will onset and terminate together.

At test, the grackles will receive four types of tests. Half of the tests involve the presentation of the Tone and the Noise on separate trials. The other half of the tests involve trials on which when the grackle pecks at one of two novel response keys (white square with gray lines (2.2cm wide x 2.2cm tall; formerly blue square) and white triangle with gray lines (3cm base x 3cm height; formerly green triangle) made available on the touchscreen, this is followed immediately by the presentation of the Tone (for pecking one response key) or the Noise (for the other response key). Thus, these latter tests involve the grackle intervening on the Tone and the Noise. If grackles have formed the causal models shown in Figure 1, then they should expect food on test trials on which the Tone and the Noise are observed (but not intervened on). This is due to the causal inferences derived from observing a direct cause (Noise) of an effect (Food), or an effect (Tone) of a direct cause (Light) of the Food. We predict different expectation of food in the Intervention test trials, however. When the grackle intervenes on the Noise, they should still expect Food since the Noise is a direct cause of Food. When the grackle intervenes on the Tone, however, they should treat the Tone as caused by their own action (key peck) and thus discount the possibility that the Light had just occurred. By discounting the Light, they should also not expect Food. Thus, we expect less Food seeking on test trials on which grackles intervene on the Tone compared to the other three trial types. If causal inference is demonstrated in Experiment 1, we will begin Experiment 2.

## Experiment 2



The materials used in Experiment 2 were identical to those used in Experiment 1. Grackles experience either common-cause training (as in Experiment 1) or causal-chain training, which is the same as the common-cause training except that the tone preceded the light during observational learning (in common cause training,  $L \rightarrow T$ ). In experiment 2, in both common-cause and causal-chain training there was no training of the noise. At test, grackles in the Intervene condition hear a tone every time a key is pecked, whereas those in the Observe condition hear a tone periodically presented in the absence of a key peck. Both the Observe and Intervene grackles in the causal chain experiment should expect food after the tone. In contrast, grackles in the common cause experiment Observe condition should expect food if they hear the tone, but not those in the Intervene condition who have only experienced the light causing the tone or the food, but not the tone causing the light or the food. Also note there was no testing of the noise in experiment 2.

### **Assignment to conditions: counterbalancing and randomization**

*Sex* is balanced across each experiment (50% female in each experiment) and allocated evenly across treatment conditions.

*Experiment 1:* Each individual will experience all four conditions in Experiment 1: 1) Intervene-Tone, 2) Intervene-Noise, 3) Observe-Tone, and 4) Observe-Noise. The Intervene condition will occur in one test session and the Observe condition in a separate test session, and the order will be counterbalanced across subjects.

*Experiment 2:* Individuals will be randomly assigned to Experiment 2a or 2b, and individuals within each experiment will receive Observe and Intervene conditions.

To prevent their previous history with causal inference experiments from confounding the results, grackles that participated in Experiment 1 will not participate in Experiment 2 - new individuals will be selected for Experiment 2.

### **Blinding of conditions during analysis**

No blinding is involved in this study.

### **Dependent variables**

- 1) The number of key pecks to the food delivery symbol (food key) on the touchscreen that releases food into the food dispenser
- 2) The number of key pecks to the novel stimuli (square and triangle) on the touchscreen

### **Independent variables**

#### ***Prediction 1: causal map***

- 1) Condition (Intervene Tone, Intervene Noise, Observe Tone, Observe Noise)
- 2) Number of trials to reverse a preference in the last reversal that individual participated in
- 3) Average latency to attempt to solve a new loci after solving a different loci (multi-access box)
- 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 2 and 3. If this ends up being the case, we will modify the code in the analysis plan below to reflect this change.
- 5) ID (random effect because multiple measures per individual)

### *Prediction 2: common-cause vs causal chain*

- 1) Condition (Intervene, Observe)
- 2) Number of trials to reverse a preference in the last reversal that individual participated in
- 3) Average latency to attempt to solve a new loci after solving a different loci (multi-access)
- 4) Flexibility comprehensive: This measure is currently being developed and is intended to 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 2 and 3. If this ends up being the case, we will modify the code in the analysis plan below to reflect this change.
- 5) 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 3.6.3; 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.

We realize that there are other variables that are not included in the analyses below that may have an influence in our models if they were included (e.g., individual differences in body size, sex, exploration, boldness, etc.). Many of these variables we will have measured on these particular individuals. We have chosen to keep the models as simple as possible because the sample sizes for each experiment are small. These experiments were designed to determine whether grackles attend to causal cues or not. If results show that they do, then we will conduct further tests to investigate the extent of these abilities. The combination of conducting multiple experiments on the same cognitive ability on different individuals at different times and locations will not only increase our overall sample size, but it will show that we were able to detect the trait we were measuring.

### *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 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). 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 analyses 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 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. The number of predictor variables was restricted to only the fixed effects because this test was not designed for mixed models. 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 ( $n=16$ ). The protocol of the power analysis applies to each of the models below because all have the same sample sizes and the same number of fixed effects (explanatory variables):

*Input:*

Effect size  $f^2 = 0,77$

err prob = 0,05

Power  $(1 - \text{err prob}) = 0,7$

Number of predictors = 3

*Output:*

Noncentrality parameter = 12,3200000

Critical F = 3,4902948

Numerator df = 3

Denominator df = 12

Total sample size = 16

Actual power = 0,7052261

This means that, with our sample size of 16 (for each experiment), we have a 71% chance of detecting a large effect (approximated at  $f^2=0.35$  by Cohen (1988)).

### **Data checking**

The data will be visually checked to determine whether they are normally distributed. Normality is indicated when the histograms of actual data match those with simulated data (Figure 2) (Zuur, Ieno, and Saveliev 2009).

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

## Check the dependent variables for normality: Histograms
op <- par(mfrow = c(2, 2), mar = c(4, 4, 2, 0.2))
# This is what the distribution of actual data looks like
hist(cause$KeyPecksFood, xlab = "Number of food key pecks", main = "Actual Data")
hist(cause$KeyPecksNovel, xlab = "Number of stimulus key pecks",
  main = "Actual Data")

# Given the actual data, this is what a normal distribution
# would look like
Y2 <- rnorm(1281, mean = mean(cause$KeyPecksFood), sd = sd(cause$KeyPecksFood))
hist(Y2, xlab = "Number of food key pecks", main = "Simulated Data")

Z2 <- rnorm(1281, mean = mean(cause$KeyPecksNovel), sd = sd(cause$KeyPecksNovel))
hist(Z2, xlab = "Number of stimulus key pecks", main = "Simulated Data")

## Check the dependent variables for normality: Q-Q plot
op <- par(mfrow = c(2, 4), mar = c(4, 4, 2, 0.2))
plot(glm(cause$KeyPecksFood ~ cause$Condition))
plot(glm(cause$KeyPecksNovel ~ cause$Condition))
```

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

```

sep = ",", stringsAsFactors = F)

op <- par(mfrow = c(1, 1), mar = c(4, 4, 2, 0.2))
plot(residuals(glm(cause$KeyPecksFood ~ cause$KeyPecksNovel)),
     ylab = "Pecks to food key ~ Pecks to novel key")

```

### *Prediction 1: causal map*

**Analysis:** 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; (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 ( $V=1$ ,  $\nu=0$ ) (Hadfield 2014). We will ensure the GLMM shows acceptable convergence (lag time autocorrelation values  $<0.01$ ; (Hadfield 2010)), and adjust parameters if necessary to meet this criterion. We will determine whether an independent variable had an effect or not using the Estimate in the full model.

```

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

# Select only data from Experiment 1
cause <- cause[cause$Experiment == 1, ]

# GLMM
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)))

# GLMM with response variable = key pecks to the food key
cause1 <- MCMCglmm(KeyPecksFood ~ Condition + AvgTrialsReverse +
  AvgLatencySwitch, random = ~ID, family = "poisson", data = cause,
  verbose = F, prior = prior, nitt = 130000, thin = 10, burnin = 30000)
summary(cause1)
# autocorr(cause1$Sol) #Did fixed effects converge?
# autocorr(cause1$VCV) #Did random effects converge?

# GLMM with response variable = key pecks to the stimulus key
cause1a <- MCMCglmm(KeyPecksNovel ~ Condition + AvgTrialsReverse +
  AvgLatencySwitch, random = ~ID, family = "poisson", data = cause,
  verbose = F, prior = prior, nitt = 130000, thin = 10, burnin = 30000)
summary(cause1a)
# autocorr(cause1a$Sol) #Did fixed effects converge?
# autocorr(cause1a$VCV) #Did random effects converge?

```

### *Prediction 2: common-cause vs causal chain*

**Analysis:** 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; (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 ( $V=1$ ,  $\nu=0$ ) (Hadfield 2014). We will ensure the GLMM shows acceptable convergence (lag time autocorrelation values  $<0.01$ ; (Hadfield 2010)), and adjust parameters if necessary to meet this criterion. We will determine whether an independent variable had an effect or not using the Estimate in the full model.

```

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

# Select only data from Experiment 2
cause <- cause[cause$Experiment == "2a" | cause$Experiment ==
  "2b", ]

# GLMM
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)))

# GLMM with response variable = key pecks to the food key
cause2 <- MCMCglmm(KeyPecksFood ~ Condition + AvgTrialsReverse +
  AvgLatencySwitch, random = ~ID, family = "poisson", data = cause,
  verbose = F, prior = prior, nitt = 13000, thin = 10, burnin = 3000)
summary(cause2)
# autocorr(cause2$Sol) #Did fixed effects converge?
# autocorr(cause2$VCV) #Did random effects converge?

# GLMM with response variable = key pecks to the stimulus key
cause2a <- MCMCglmm(KeyPecksNovel ~ Condition + AvgTrialsReverse +
  AvgLatencySwitch, random = ~ID, family = "poisson", data = cause,
  verbose = F, prior = prior, nitt = 13000, thin = 10, burnin = 3000)
summary(cause2a)
# autocorr(cause2a$Sol) #Did fixed effects converge?
# autocorr(cause2a$VCV) #Did random effects converge?

```

### Alternative Analyses

Logan anticipates that she 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], SP606267 [2018], and SP639866 [2019])
- 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 [2017])

## G. AUTHOR CONTRIBUTIONS

**Blaisdell:** Hypothesis development, experimental design, data analysis and interpretation, write up.

**Johnson-Ulrich:** Touchscreen programming, data collection, data interpretation, revising/editing.

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

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

**Seitz:** Programmed the touchscreen experiment, revising/editing.

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

**Logan:** Hypothesis development, data collection, data analysis and interpretation, write up, 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, in 2017 through early 2018, by a Leverhulme Early Career Research Fellowship to Logan.

## I. ACKNOWLEDGEMENTS

We thank Dieter Lukas for help polishing the hypotheses; Ben Trumble for hosting the grackle project at Arizona State University (providing office and lab space); 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 two anonymous reviewers and the Recommender, Emanuel Fronhofer, at PCI Ecology for their useful feedback.

## J. REFERENCES

- Blaisdell, Aaron P, Kosuke Sawa, Kenneth J Leising, and Michael R Waldmann. 2006. "Causal Reasoning in Rats." *Science* 311 (5763). American Association for the Advancement of Science: 1020–2.
- 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.
- Griffin, Andrea S, and David Guez. 2014. "Innovation and Problem Solving: A Review of Common Mechanisms." *Behavioural Processes* 109. Elsevier: 121–34.
- Hadfield, JD. 2010. "MCMC Methods for Multi-Response Generalized Linear Mixed Models: The Mcmcglmm R Package." *Journal of Statistical Software* 33 (2): 1–22.
- . 2014. "MCMCglmm Course Notes." <http://cran.r-project.org/web/packages/MCMCglmm/vignettes/CourseNotes.pdf>.
- Lefebvre, Louis, Patrick Whittle, Evan Lascaris, and Adam Finkelstein. 1997. "Feeding Innovations and Forebrain Size in Birds." *Animal Behaviour* 53 (3). Elsevier: 549–60.

- Leising, Kenneth J, Jared Wong, Michael R Waldmann, and Aaron P Blaisdell. 2008. "The Special Status of Actions in Causal Reasoning in Rats." *Journal of Experimental Psychology-General* 137 (3). [Washington] American Psychological Association.: 514–27.
- 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.
- Peirce, Jonathan W. 2009. "Generating Stimuli for Neuroscience Using Psychopy." *Frontiers in Neuroinformatics* 2. Frontiers: 10.
- R Core Team. 2017. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org>.
- 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, Tamas Székely, Andras Liker, and Louis Lefebvre. 2007. "Big-Brained Birds Survive Better in Nature." *Proceedings of the Royal Society of London B: Biological Sciences* 274 (1611). The Royal Society: 763–69.
- 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.
- Bateson, Melissa, Michael Emmerson, Gökçe Ergün, Pat Monaghan, and Daniel Nettle. 2015. "Opposite Effects of Early-Life Competition and Developmental Telomere Attrition on Cognitive Biases in Juvenile European Starlings." *PLoS One* 10 (7). Public Library of Science: e0132602.
- Blaisdell, Aaron P, Kosuke Sawa, Kenneth J Leising, and Michael R Waldmann. 2006. "Causal Reasoning in Rats." *Science* 311 (5763). American Association for the Advancement of Science: 1020–2.
- 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.
- Griffin, Andrea S, and David Guez. 2014. "Innovation and Problem Solving: A Review of Common Mechanisms." *Behavioural Processes* 109. Elsevier: 121–34.

- Hadfield, JD. 2010. “MCMC Methods for Multi-Response Generalized Linear Mixed Models: The Mcmcglmm R Package.” *Journal of Statistical Software* 33 (2): 1–22.
- . 2014. “MCMCglmm Course Notes.” <http://cran.r-project.org/web/packages/MCMCglmm/vignettes/CourseNotes.pdf>.
- Lefebvre, Louis, Patrick Whittle, Evan Lascaris, and Adam Finkelstein. 1997. “Feeding Innovations and Forebrain Size in Birds.” *Animal Behaviour* 53 (3). Elsevier: 549–60.
- Leising, Kenneth J, Jared Wong, Michael R Waldmann, and Aaron P Blaisdell. 2008. “The Special Status of Actions in Causal Reasoning in Rats.” *Journal of Experimental Psychology-General* 137 (3). [Washington] American Psychological Association.: 514–27.
- 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. <https://doi.org/10.1098/rsfs.2016.0121>.
- Peer, Brian D. 2011. “Invasion of the Emperor’s Grackle.” *Ardeola* 58 (2). BioOne: 405–9.
- Peirce, Jonathan W. 2009. “Generating Stimuli for Neuroscience Using Psychopy.” *Frontiers in Neuroinformatics* 2. Frontiers: 10.
- R Core Team. 2017. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <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). National Acad Sciences: 5460–5.
- 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, Sarah Timmermans, and Louis Lefebvre. 2002. “Behavioural Flexibility and Invasion Success in Birds.” *Animal Behaviour* 63 (3). Elsevier: 495–502.
- Waldmann, Michael R, and York Hagmayer. 2005. “Seeing Versus Doing: Two Modes of Accessing Causal Knowledge.” *Journal of Experimental Psychology: Learning, Memory, and Cognition* 31 (2). American Psychological Association: 216.
- 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.