Innovation does not indicate behavioral flexibility in great-tailed grackles

Corina J. Logan*

SAGE Center for the Study of the Mind, University of California Santa Barbara, Santa Barbara, CA 93103 USA

*Current address: Department of Zoology, University of Cambridge, Cambridge CB2 3EJ United Kingdom, e-mail: cl417@cam.ac.uk

Note: this manuscript was submitted to Animal Cognition.

ABSTRACT

Many cross-species studies attest that innovation frequency (novel food types eaten and foraging techniques used) is a measure of behavioral flexibility and show that it positively correlates with relative brain size (corrected for body size). However, mixed results from the three studies that directly test the relationship between innovation frequency and behavioral flexibility and behavioral flexibility and brain size question both assumptions. I investigated behavioral flexibility in non-innovative great-tailed grackles that have an average sized brain, and compared their test performance with innovative, large-brained New Caledonian crows. Contrary to the prediction, grackles perform similarly to crows in experiments using clear tubes partially filled with water and containing a floating food reward, where objects must be dropped into the tube to raise the water level, bringing the food within reach. Similarly to crows, grackles preferred to drop the more functional heavy (rather than light) objects and some changed their preference in a follow up experiment where the heavy objects were no longer functional, thus exhibiting behavioral flexibility. These results challenge the assumption that innovation frequency indicates behavioral flexibility since a non-innovative bird demonstrated behavioral flexibility at a level similar to that in innovative crows, and they challenge the assumption that only large brains are capable of behavioral flexibility because a bird with an average brain size solved problems similarly to large-brained crows.

KEYWORDS

Innovation, behavioral flexibility, brain size, birds, water displacement, Aesop's Fable, comparative cognition

INTRODUCTION

Behavioral flexibility, currently defined as the ability to adapt behavior to changing contexts, is considered the keystone of complex cognition (Buckner 2013). Measuring behavioral flexibility directly in each species is time intensive. Thus, comparative biologists seek behaviors that can serve as indicators of behavioral flexibility, therefore allowing cross-species comparisons of cognition with behavior, ecology, and life history (Lefebvre et al. 1997; see Healy and Rowe 2007 for a review). One widely used indicator of behavioral flexibility is the frequency of innovations, where innovations are based on reports of novel food types eaten and foraging techniques used (Lefebvre et al. 1997, 2004, 2013; Timmermans et al. 2000; Nicolakakis and Lefebvre 2000; Reader and Laland 2002). This operational definition of behavioral flexibility relies on two main assumptions: 1) innovativeness indicates complex cognition through behavioral flexibility and 2) innovativeness actually measures behavioral flexibility (e.g., Lefebvre et al. 1997, 2002; Timmermans et al. 2000). The first assumption presumes that behavioral flexibility can serve as evidence of complex cognition because those individuals with more cognitive processing power (as indicated by relative brain size [corrected for body size]) should be able to adapt their behavior more flexibly to changing circumstances. However, this is a circular argument that needs validation from external factors. Without this validation, progress can only be made on the second assumption. Here, I review the evidence provided by direct tests of behavioral flexibility in species that vary in their innovation frequencies and relative brain sizes, and I test the assumption that innovation frequency is a proxy for behavioral flexibility by comparing two bird species that differ in both respects.

Results from experiments testing behavioral flexibility show mixed evidence from cross-species studies about how behavioral flexibility relates to innovation frequency and relative brain size. There is some evidence that behavioral flexibility correlates with innovation frequency, but not relative brain size: Innovative, smaller-brained Galapagos finches reversed a previously learned color preference faster than less innovative, relatively larger-brained New World jays (Tebbich et al. 2010). Other evidence shows that behavioral flexibility correlates with relative brain size, but not innovation frequency: Keas and New Caledonian crows performed similarly on a multi-access box — where the successful solution of an option resulted in its closure, thereby forcing the individual to innovate another solution on the same box — even though the crows are innovative and the keas are not (Auersperg et al. 2011). Evidence from primates shows that behaviorally flexible problem solving does not correlate with innovation frequency or relative brain size: Chimpanzees, bonobos, and gorillas performed better on a multi-access box than orang-utans, and all species (except for all but one orang-utan) quickly moved on to other techniques for accessing the food when the current method stopped working (Manrique et al. 2013). All of these primate species are innovative and have relatively large brains, with gorillas having the smallest of this group of species (Isler et al. 2008). The multi-access box experiment measures behavioral flexibility in successful individuals by requiring them to adapt their behavior to changing circumstances. Therefore, so far, direct evidence indicates that innovation frequency may not measure behavioral flexibility and that behavioral flexibility does not correlate with relative brain size. The one study that found evidence supporting the link between innovation frequency and behavioral flexibility comes from the only study to compare species with different brain sizes (Tebbich et al. 2010). The other two studies examined relatively large brained species and found a consistent lack of support regarding the link between innovation frequency and behavioral flexibility. Therefore, it is unclear how innovation frequency relates to brain size when examining behavioral flexibility directly.

The aim of my study was to test the relationship between innovation frequency and behavioral flexibility, and behavioral flexibility and relative brain size in species with differing brain sizes by directly measuring behavioral flexibility in great-tailed grackles (*Quiscalus mexicanus*) and comparing their performance with previously tested New Caledonian crows (*Corvus moneduloides*; Logan et al. 2014). Grackles are not innovative (n=1 innovation; Ducatez et al. 2014) and have an average relative brain size, while New Caledonian crows are innovative (n=8 innovations; literature reviewed by Logan: Layard & Layard 1882, Le Goupils 1928, Hunt 1996, 2008, Hunt & Gray 2004) and have relatively large brains (relative brain size data for 2131 bird species from: Alma and Bee de Speroni 1992; Bee de Speroni and Carezzano 1995; Boire and Baron 1994; Carezzano et al. 1995; Crile and Quiring 1940; Day et al. 2005; Ebinger 1995; Ebinger and Lohmer 1984, 1987; Ebinger and Rohrs 1995; Fernandez et al. 1997; Frahm and Rehkamper 1998, 2004; Iwaniuk and Nelson 2003; Iwaniuk 2004; Iwaniuk et al. 2004, 2005; Iwaniuk and Wylie 2006; Iwaniuk unpublished data [380 species]; Milkovsky 1989a,b,c, 1990; Møller et al. 2004; Pistone et al. 2002; Rehkamper et al. 1991, 2003, 2008). Therefore, the prediction is that the less innovative and smaller-brained grackles will show less behavioral flexibility than crows.

I tested behavioral flexibility in grackles using the water tube paradigm (or Aesop's Fable paradigm), which has previously been used to explore the cognitive abilities that underlie problem solving (Bird & Emery 2009; Cheke et al. 2011, 2012; Taylor et al. 2011; Jelbert et al. 2014; Logan et al. 2014). This research has shown that corvids (birds in the crow family) prefer to drop heavy objects that sink, rather than light objects that float, into a water tube to raise the water level and bring floating food within reach (Cheke et al. 2011; Taylor et al. 2011; Jelbert et al. 2014; Logan et al. 2014). In these experiments, the heavy objects displaced more water than the light objects, thus raising the water level in the tube by a larger amount and bringing the food closer to the top of the tube. Previous heavy vs. light experiments (also called sinking vs. floating) used objects where the heavy items (rubber) were sinkable, but the light items (foam or polystyrene) were not, thus one needed to discriminate between discrete kinds of functionality to solve the task.

In this study, I modified the water tube experiments to investigate behavioral flexibility. I tested behavioral flexibility, the ability to change preferences when the context changes (Buckner

2013), by presenting the grackles first with the heavy vs. light experiment and then with a follow up experiment in which the heavy objects were no longer functional. In this follow up experiment (heavy vs. light magic), heavy objects stuck to a magnet placed inside the tube above the water level, leaving the light objects as the functional option because they could fall past the magnet and into the water. If grackles preferred heavy objects or had no preference in the heavy vs. light experiment and then changed their preference in the heavy vs. light magic experiment to preferring neither object or light objects, this would indicate that their preferences are sensitive to changing contexts. New Caledonian crows exhibited behavioral flexibility using the water tube tests when they discriminated between two tubes of different volumes (Logan et al. 2014). In the first experiment, crows preferred to drop objects into a narrow (functional) rather than a wide (non-functional) tube when water levels were equal in both tubes. In a follow up experiment where the narrow tube was no longer functional because the water level was too low, crows changed their preference to dropping objects into the functional wide tube. I carried out these same experiments with the grackles to compare their flexibility with that in New Caledonian crows.

To summarize, behavioral flexibility would be shown if the grackles that preferred heavy in heavy vs. light changed their preference to no preference or to preferring light objects in heavy vs. light magic experiment, and if those grackles that preferred the narrow tube in narrow vs. wide with equal water levels experiment changed their preference to the wide tube in narrow vs. wide with unequal water levels experiment. The crows were not given the heavy vs. light magic experiment because it had not been designed yet, therefore grackle and crow behavioral flexibility could be directly compared using the wide vs. narrow equal and unequal water level experiments. Behavioral flexibility in these two species could be more generally compared in terms of their ability to change preferences when circumstances change regardless of which experiments they demonstrate flexibility in.

METHODS

Ethics

This research was carried out in accordance with permits from the U.S. Fish and Wildlife Service (scientific collecting permit number MB76700A), California Department of Fish and Wildlife (scientific collecting permit number SC-12306), U.S. Geological Survey Bird Banding Laboratory (federal bird banding permit number 23872), and the Institutional Animal Care and Use Committee at the University of California Santa Barbara (IACUC protocol number 860).

Subjects and Study Site

Eight wild adult great-tailed grackles (4 females and 4 males) were caught using a walk-in baited trap measuring 2ft high by 2ft wide by 4ft long (design from Overington et al. 2011). Birds were caught (and tested) in two batches: batch one at the Andree Clark Bird Refuge (4 birds in September 2014, released in December) and batch two at East Beach Park (4 birds in January 2015, released in March) in Santa Barbara, California. They were housed individually in aviaries measuring 72in high by 47in wide by 93in long at the University of California Santa Barbara for 2-3 months while participating in the experiments in this study. Grackles were given water *ad libitum* and unrestricted amounts of food (Mazuri Small Bird Food, bread, and peanuts) for at least 20 hrs per day, with their main diet being removed for up to 4 hrs on testing days while they participated in experiments. Grackles were aged by plumage and eye color and sexed by plumage and weight following Pyle (2001). Biometrics, blood, and feathers were collected at the beginning and end of their time in the aviary. Their weights were measured at least once per month, first at the time of trapping using a balancing scale, and subsequently by placing a kitchen scale covered with food in their aviary and recording their weight when they jumped onto the scale to eat.

Color Learning to Prevent Side Bias

To help break potential side biases during the wide vs. narrow water tube experiment, I first had grackles learn a simple association between food and color, which forced them to pay attention to color rather than spatial location (see Logan et al. 2014). They were given a silver and a gold tube with food always hidden in the gold tube. One silver and one gold tube were placed at opposite ends of a table with the tube openings facing the side walls so the bird could not see which tube contained the food. Tubes were pseudorandomized for side and the left tube was always placed on the table first, followed by the right to avoid behavioral cueing. Pseudorandomization involved alternating sides for the first two trials in a 10-trial set and then never having one tube on the same side for more than two trials in a row, while avoiding a pattern that would allow the bird to follow a rule to solve the task rather than learning which color indicated the food. Each trial consisted of placing the tubes on the table and then the bird had the opportunity to choose one tube by looking into it (and eating from it if it chose the gold tube). Once the bird chose, the trial ended by interrupting the bird and removing the tubes. A bird passed this test if it made at least 17 correct choices out of the most recent 20 trials. Proficiency with this test then served as a useful tool for later water tube experiments involving two tubes: if a grackle developed a side bias, the water tube experiment was paused and silver/gold tests were conducted until the bird attended to color rather than location (side).

Spontaneous Stone Dropping

Birds were given two sequential 5 minute trials with the stone dropping training apparatus and two stones to see whether they would spontaneously drop stones down tubes. The stone dropping training apparatus was a clear acrylic box with a tube on top. The box contained out of reach food on top of a platform that was obtainable by dropping a stone into the top of the tube, which, when contacting the platform, forced the magnet holding it up to release the platform (design as in Bird and Emery 2009 with the following tube dimensions: 90mm tall, outer diameter=50mm, inner diameter=37 or 44mm; Figure 1). The food then fell from the platform to the table. At the end of the first 5 minute trial, the stones were moved to different locations on the table and on the wooden blocks. The blocks made it easier to access the top of the tube.

Figure 1. Batido participates in stone dropping training.

Stone Dropping Training

Those birds that did not spontaneously drop stones down the tube on the stone dropping training apparatus were trained to push or drop stones down tubes using this apparatus. Birds were given two stones and went from accidentally dropping stones down the tube as they pulled at food under the stones, which were balanced on the edge of the tube opening, to pushing or dropping stones into the tube from anywhere near the apparatus. Once the bird proficiently pushed or dropped stones into the apparatus 30 times, they moved onto the reachable distance on a water tube. Stone pushing/dropping proficiency was defined as consistently directing the stone to tube opening from anywhere on the ramp on the top of the apparatus. Not all motions had to be in the direction of the tube opening because some grackles preferred to move the stone to a particular location on the ramp (which may initially be in the opposite direction from the tube) and push or drop it in from there or push the stone in shorter, angular strokes. It was permissible for a bird to throw one of the stones off the side of the apparatus (which occurred sporadically throughout all of their experiences with stone pushing/dropping) as long as they proficiently put the other stone in the tube.

Reachable Distance

To determine how high to set the water levels in water displacement experiments, a bird's reachable distance was obtained. Food was placed on cotton inside a resealable plastic bag, which was stuffed inside the standard water tube (a clear acrylic tube [170mm tall, outer diameter=51mm, inner diameter=38mm] super glued to a clear acrylic base [300x300x3mm]) to obtain the reachable distance without giving the bird experience with water (Figure 2). The food was first placed within reach and then lowered into the tube in 1cm increments until the bird could not reach it. The lowest height the bird could still reach was considered its reachable distance and water levels in subsequent experiments were set to allow the desired number of objects to bring the food within reach.

Figure 2. Obtaining Tequila's reachable distance by placing food on top of cotton wrapped in plastic to avoid giving him experience with water before the experiments.

Water Tube Proficiency Assessment

To determine whether individuals transfer their stone pushing/dropping skills from a tube on a platform to a tube containing water or whether they need additional training on this new apparatus, they were given a partially filled water tube with a floating peanut piece and four stones (9-14g, each displaces 5-6mm water) which they could drop into the tube to raise the water level and consequently reach the food (Figure 3). Once a bird accomplished 30 consecutive proficient trials, they moved onto experiment 1. Proficiency was defined as in the stone dropping training section above.

Figure 3. Batido participates in the water tube proficiency assessment.

Experimental Set Up

Apparatuses were placed on top of rolling tables (23.5in wide by 15.5in long) and rolled into each individual's aviary for testing sessions, which lasted up to approximately 20min. Tubes were baited with 1/16 of a peanut attached to a small piece of cork with a tie wrap for buoyancy (peanut float). The area around the top of the tube next to the objects available for dropping in the tube was also sometimes baited with smaller peanut pieces and bread crumbs to encourage the bird to interact with the task. All experiments consisted of 20 trials per bird.

Experiment 1: Heavy vs. Light

A water tube was presented with 4 heavy (steel rod wrapped in fimo clay, weight=10g, each displaces 2-3mm of water) and 4 light (plastic tube partially filled with fimo clay, weight=2g, each displaces 1-1.5mm of water) objects placed in pseudorandomized (as explained for color learning) pairs near the top of the tube (both objects were 21-24mm long and 8mm in diameter; Figure 4). Heavy objects displaced 0.5-2mm more water than light objects, thus making them more functional than the light objects, but importantly, both objects were functional.

Figure 4. The Heavy vs. Light experimental set up.

Experiment 2: Heavy vs. Light Magic

The set up was the same as in experiment 1, except there were magnets (2 super magnets on the outside and 3 inside of the tube) attached to the tube above the water level such that the heavy objects would stick to the magnets and not displace water, while the light objects could fall past the magnets into the water, thus being the functional choice (Figure 5). Birds were given 3 heavy and 3 light objects, placed in pseudorandomized pairs near the top of the tube.

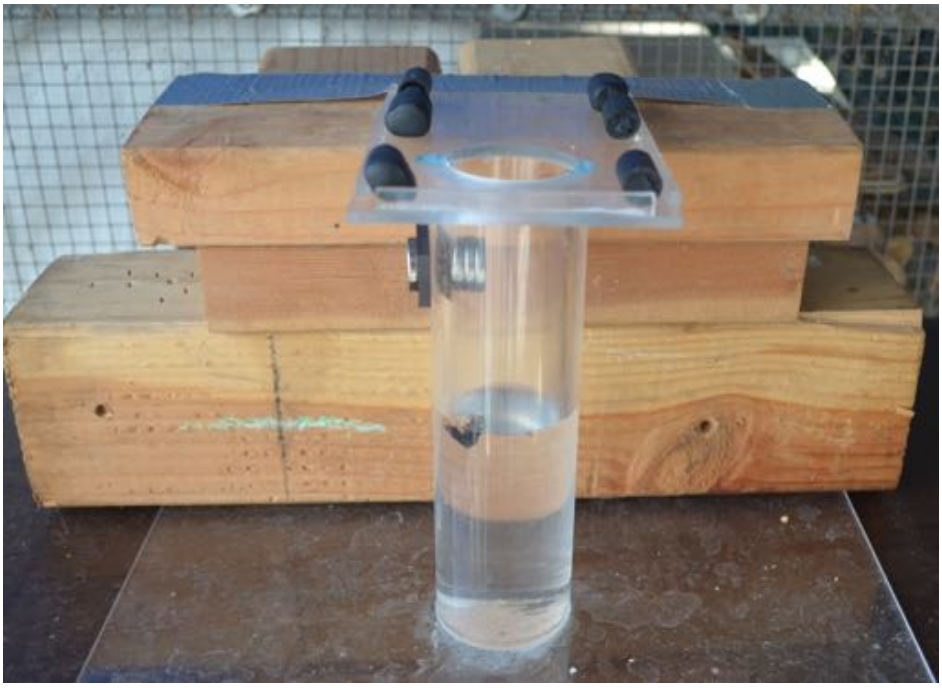

Figure 5. The Heavy vs. Light Magic experimental set up, which includes magnets stuck to the tube above the water level.

Experiment 3: Narrow vs. Wide Equal Water Levels

To determine whether birds understand volume differences, a wide and narrow tube with equal water levels were presented with four objects made out of fimo clay (30x10x5mm, 3-4g, each object displaced 1-2mm in wide tube and 5-6mm in narrow; Logan et al. 2014; Figure 6). Two objects were placed near the narrow tube opening and two objects near the wide tube opening. The objects were only functional if dropped into the narrow tube because the water levels were set such that dropping all of the objects into the wide tube would not bring the floating food within reach. However, dropping 1-2 objects into the narrow tube would raise the water level enough to reach the food. Both tubes were 170mm tall with 3mm thick lids that constricted the opening to 25mm in diameter to equalize the bird's access to the inside of each tube, and super glued to a clear acrylic base (300x300x3mm). The wide tube (outer diameter=57mm, inner diameter=48mm, volume=307,625mm^3) was roughly equally larger than the standard water tube (dimensions above, volume=192,800mm^3) as the narrow tube was smaller (outer diameter=38mm, inner diameter=25mm, volume=83,449mm^3). The position of the tubes was pseudorandomized for side to ensure that tube choices were not based on a side bias.

Experiment 4: Narrow vs. Wide Unequal Water Levels

Those grackles that passed experiment 3 continued to this experiment to determine whether their tube choices adjusted to changing circumstances. This experiment was the same as experiment 3, except the water level in the narrow tube was lowered to 5cm from the table, thus making the food unreachable even if all objects were dropped into this tube (as in Logan et al. 2014). The water level in the wide tube was raised such that the bird could reach the food in 1-4 object drops.

Figure 6. The Narrow vs. Wide with equal water levels experimental set up.

Experimenters

I conducted all experiments, and my students (Luisa Bergeron, Alexis Breen, Michelle Gertsvolf, Christin Palmstrom, and Linnea Palmstrom) and I conducted the stone dropping training.

Statistical Analyses

To make this research comparable with previous studies, I used binomial tests to determine whether each grackle chose particular objects or tubes at random chance (null hypothesis: $p \geq 0.05$) or significantly above chance (alternative hypothesis: $p < 0.05$). The Bonferroni-Holm correction was applied to p-values within each experiment to correct for an increase in false positive results that could arise from conducting multiple tests on the same dataset.

Generalized linear mixed models (GLMMs) were used to determine whether birds preferred particular objects or tubes (response variable: correct or incorrect choice) in an experiment and whether the trial number or bird influenced choices (explanatory variables: experiment, trial number, bird), and to control for the non-independence of multiple choices per trial (random factor: choice number). I set the prior to fix the variance component to one (fix=1) because the measurement error variance was known, as is standard when choices are binary (Hadfield 2010). I ensured that the Markov chain for this test model converged by manipulating the number of iterations (nitt=150000), the number of iterations that must occur before samples are stored (burnin=30000), and the intervals the Markov chain stores (thin=300) until successive samples were independent (autocorr function, MCMCglmm package: Hadfield 2010) and there were no trends when visually inspecting the time series of the Markov chain (function: plot(testmodel$Sol); Hadfield 2014). I compared this test model to a null model where I removed all explanatory factors and set it to 1. I determined whether the test model was likely given the data, relative to the null model by using Akaike weights (range: 0-1, all model weights sum to 1; Akaike 1981; Weights function, MuMIn package: Bates et al. 2011). The Akaike weight indicates the “relative likelihood of the model given the data” (Burnham and Anderson 2002, p. xxiii) and models with Akaike weights greater than 0.9 are considered reliable models because they are highly likely given the data (Burnham and Anderson 2002). The test model was highly likely given the data (Akaike weight=0.99) and the null model was not (Akaike weight=0.0009). To investigate potential effects of season or order of testing, I carried out a GLMM to investigate whether the batch to which the bird belonged (explanatory variable: batch=1 or 2) influenced their test performance (response variable: correct or incorrect choice) while controlling for the non-independence of multiple choices per trial (random factor: choice number). The null model was highly likely given the data (Akaike weight=0.94), while the batch model was not (Akaike weight=0.06), indicating that batch did not influence test performance. GLMMs were carried out in R v3.1.2 (R Core Team 2014) using the MCMCglmm function (MCMCglmm package) with a binomial distribution (called categorical in MCMCglmm) and logit link.

Data Availability

The data used for the GLMM, including each choice for every bird in all experiments, is available at the KNB Data Repository: https://knb.ecoinformatics.org/#view/corina_logan.15.4 (Logan 2015).

RESULTS

Spontaneous Stone Dropping

No grackle spontaneously dropped stones down the tube of the platform apparatus. Therefore, they all underwent stone dropping training.

Stone Dropping Training

Most grackles learned to push stones into a tube on the platform apparatus in 165-392 trials (Table 1), however Michelada was scared of the stone falling down the tube and did not habituate to this event and Jugo learned too slowly to become proficient by the time he needed to be released, therefore they were excluded from the stone dropping experiments. The training procedure was modified from Logan et al. (2014) to allow stone pushing from a clear cast acrylic ramp placed on top of the tube rather than stone dropping by picking up the stone from the table and putting it into the tube (without a ramp; see Figures 1-6). The modification was necessary because grackles seem to form associations between the stones and the top of the tube, the stones and the table where the food comes out, and the stones falling only in one direction: down. When I placed the stones below the level of the top of the tube to try to train them to pick the stones up and put them in the top of the tube, the grackles took the stones and dropped them off the side of the apparatus or table, often placing them on the table and then looking at where the platform should have fallen open, awaiting the food. Placing the ramp on the water tubes for the experiments was implemented to mitigate this limitation. Once this change was made, it was no longer necessary to train the grackles to pick up and drop the stones because pushing them into the tube sufficed and required less training.

Water Tube Proficiency Assessment

Most grackles immediately applied their stone dropping skills to a water tube context as indicated by their first 30 trials being proficient (Cerveza, Margarita, Refresco, Batido). Horchata took 31 trials to reach proficiency. Tequila did initially apply his stone dropping skills to a water tube context, however his order of experiments was different: he went from determining his reachable distance to an experiment involving a water-filled and a sand-filled tube, filled to equal levels. He participated in three trials, but lost motivation and started to give up on participating in stone dropping all together. The water tube proficiency assessment was then developed to remotivate him to participate in subsequent experiments, and the sand vs. water experiment was eliminated. After this additional experience, Tequila needed 106 trials to reach the water tube proficiency criteria.

Experiment 1: Heavy vs. Light

Four grackles (Tequila, Margarita, Batido, and Refresco) were 3.2-4.9 times more likely to choose heavy objects rather than the less functional light objects, while two grackles (Cerveza and Horchata) had no preference (they were 0.6-1.4 times more likely to succeed than fail; see Table 1 for binomial test results and Table 2 for GLMM results, Supplementary Material Video 1 available at: https://youtu.be/Wa44bz9MU_8). Cerveza and Horchata's performances improved across trials: they were 3.6-4.1 times more likely to succeed than fail as trial number increased, indicating that they learned through trial and error that the heavy objects were more functional (Table 2). The other grackles' performances did not improve with increasing trial number, indicating that they used prior knowledge to solve the task (Table 2). Horchata was not motivated to participate in the water tube experiments: she required bait between almost all trials to get her to continue to interact with the apparatus, which might have influenced her lack of success. All choices in all trials for all birds in all experiments are presented in Figures S1.1-1.3 in Supplementary Material.

Table 1. Performance per bird per experiment: the number of stone dropping training trials needed to reach proficiency, and p-values from Bonferroni-Holm corrected (within experiment) binomial tests for each experiment (- = was not given this experiment). Note: Tequila was the first bird tested and I did not realize until after I trained him to pick up and drop the stones into the tube that I wanted to only train the other birds to push the stones into the tube to save training time. Therefore, the trial numbers for the other birds refer to proficiency to push objects into the tube, not pick up and drop them. Y=yellow, P=purple, B=blue, O=orange, R=red, G=green.

Bird (color rings)	**Sex**	**Stone drop training trials**	**Heavy vs. Light**	**Heavy vs. Light Magic**	**Wide vs. Narrow Equal**
Tequila (YP)	M	222 push / 263 drop	0.003 heavy	0.60	-
Margarita (PB)	F	392	0.00001 heavy	0.02 heavy	-
Cerveza (BO)	F	282	0.06	0.02 heavy	1.00
Michelada (OR)	F	-	-	-	-
Batido (OP)	M	209	0.002 heavy	0.02 heavy	1.00
Horchata (GR)	F	165	0.60	1.00	-
Refresco (PY)	M	234	0.009 heavy	1.00	1.00
Jugo (RB)	M	-	-	-	-

Table 2. Examining the influence of experiment, trial, and bird on test success (Test Performance) and whether success increased with trial number (Learning Effects), thus indicating a learning effect. GLMM: Choices Correct ~ Experiment*Trial*Bird, random = ~Choice Number. CI=credible intervals, italics indicates the intercept.

	Test Performance			**Learning Effects**		
	Posterior Mean	**Lower 95% CI**	**Upper 95% CI**	**Posterior Mean**	**Lower 95% CI**	**Upper 95% CI**
Choice number	0.002	1.39E-16	0.002	-	-	-
Heavy vs. Light						
Batido	*1.27*	*0.08*	*2.65*	-0.01	-0.13	0.13
Margarita	0.32	-1.95	2.69	0.05	-0.14	0.27
Cerveza	-1.78	-3.66	0.19	0.13	0.05	0.29
Horchata	-0.96	-3.02	0.93	0.01	-0.20	0.18
Refresco	-0.11	-1.99	2.00	-0.01	-0.17	0.16
Tequila	0.29	-1.68	2.36	-0.01	-0.18	0.21
Heavy vs. Light Magic						
Batido	-1.95	-4.25	0.19	-0.07	-0.27	0.18
Margarita	-0.35	-3.61	3.16	-0.01	-0.26	0.32
Cerveza	1.87	-1.29	4.52	-0.16	-0.46	0.11
Horchata	2.58	-0.33	6.12	-0.03	-0.34	0.25
Refresco	-1.45	-4.80	1.84	0.34	0.02	0.63
Tequila	-1.24	-4.26	2.37	0.20	-0.10	0.55
Narrow vs. Wide						
Batido	-1.03	-3.01	0.60	0.01	-0.15	0.18
Cerveza	-0.30	-3.10	2.38	0.06	-0.18	0.30
Refresco	-0.51	-3.37	2.11	0.02	-0.19	0.26

Experiment 2: Heavy vs. Light Magic

Tequila and Refresco changed from preferring heavy objects in Experiment 1 to having no preference in this experiment, while Batido continued to prefer the non-functional heavy objects (see Table 1 for binomial test results and Table 2 for GLMM results). Margarita continued to prefer heavy items and Cerveza went from having no preference to preferring the non-functional heavy items, likely due to their interest in the magnet (Table 1). The magnet seemed to attract their interest, thus continuing or increasing their preference for heavy objects (Supplementary Material Video 2 available at: https://youtu.be/TrKWEch1Y5M). Tequila gave up after 17 trials, refusing to drop either type of object into the tube, indicating he may have inhibited his choice of heavy. Tequila and Refresco's performance improved with trial number, indicating that they learned through trial and error about which object was functional (Table 2). The other grackles performances decreased with increasing trial number, indicating that they did not learn about which object was functional (Table 2). Even though Tequila and Refresco did not learn to prefer light in the amount of trials given, they did exhibit flexibility in that they changed their preferences from heavy in the previous experiment to having no preference in this experiment. Indeed, Refresco would likely have shown a preference for light objects if given more trials since all choices in his last five trials were light objects (Supplementary Material Figure S1.2).

Experiment 3: Narrow vs. Wide Equal Water Levels

All three grackles that participated in this experiment displayed no preference for dropping objects into the functional narrow tube or the non-functional wide tube (see Table 1 for binomial test results and Table 2 for GLMM results, Supplementary Material Video 3 available at: https://youtu.be/25Dj3vnSz5M). None of the grackles' performances improved with trial number, indicating that they did not learn to distinguish which one was functional (Table 2). Batido appeared to rely on the strategy of dropping all objects into both tubes regardless of which tube he received a reward from, although in trial 12, he picked up the objects from the wide tube area and dropped them into the narrow tube even though he was only trained to push stones, not drop them (Supplementary Material Figure S1.3). Since no grackle passed this experiment, they were not given experiment 4 (Narrow vs. Wide with unequal water levels), which would have investigated their behavioral flexibility in this context.

Some grackles did not initially transfer from dropping previous object types to dropping the clay objects used in this experiment. It appeared as though they were trying to solve the problem, but did not perceive the clay objects as being the kind of thing one would drop into a water tube. In these cases, additional training was implemented using a single standard water tube and a mixture of clay objects and stones until the bird was willing to drop objects into the tube even if they only consisted of clay objects. Cerveza transferred to dropping clay objects after 4 training trials, but Tequila and Margarita were excluded from this experiment because they did not transfer to dropping clay objects into tubes. After 14 training trials on a regular water tube with stones and clay objects available to Tequila, it was clear that it would take many more training trials than there was time for and his motivation was greatly diminished. Margarita refused to participate in the training trials. Horchata was also excluded from this experiment because she refused to interact with the objects.

First Choices on First Trials

All six grackles chose the more functional heavy objects as their first choice in their first trial in Heavy vs. Light, which indicates that they preferred the heavy objects from the very beginning of the experiment (Figure S1.1). Five out of six grackles chose the non-functional heavy objects in Heavy vs. Light Magic (Figure S1.2), which is not surprising given that they had learned to prefer heavy objects in the previous experiment and had likely never interacted with a magnet before, therefore they should have had no reason to have a prior understanding of how the Magic experiment worked. Two out of three grackles chose the functional narrow tube in Narrow vs. Wide with equal water levels, indicating no initial preference for a particular tube (Figure S1.3).

Did choice number influence the results?

Individuals could learn how the task worked with each choice they made, potentially making each choice dependent on previous choices. Multiple choices could be made per trial; therefore I analyzed how independent choice number was. Choice number was modeled as a random factor in the GLMM and did not influence the results, indicating that choices appear independent of each other (Table 2).

DISCUSSION

Despite their average brain size and lack of innovations, great-tailed grackles performed similarly to innovative and large-brained New Caledonian crows (Logan et al. 2014) on the Heavy vs. Light experiment, and grackles exhibited behavioral flexibility by changing their preferences in the Heavy vs. Light Magic experiment. Grackles and crows exhibited behavioral flexibility in different two-step experiments making it difficult to directly compare how similar their behavioral flexibility is: crows were not given Heavy vs. Light Magic because it was not invented yet and grackles were not able to be given Narrow vs. Wide with unequal water level experiment because no grackle passed the equal water level precursor. However, the fact that both species exhibited behavioral flexibility using the water tube paradigm allows for a more general comparison of behavioral flexibility as it relates to a species innovation frequency and brain size.

Both grackles and crows preferred to drop the functional heavy objects into the water tube rather than the less functional (for grackles) or non-functional (for crows) light objects. All grackles were successful at obtaining the food, and 4 out of 6 grackles preferred to drop the more functional heavy objects. Similar to the crows (Logan et al. 2014), Tequila, Batido, and Refresco preferred heavy objects significantly more than light objects and did not show a learning effect across the 20 trials in this experiment, indicating that they relied on prior information about the world to solve this task.

Behavioral flexibility was exhibited by grackles insofar as they changed their preferences when the task changed. When the heavy objects in the Heavy vs. Light Magic experiment were no longer functional because they stuck to a magnet, some grackles changed from having preferred heavy objects when they were functional in the previous experiment to having no object preference in the Magic experiment. This demonstrates attention to the functional properties of objects in changing circumstances. No grackle completely switched their preference to the light objects, which may have been made difficult by the design of the apparatus: if one heavy item was dropped into the tube, it stuck to the magnet and blocked access to the food regardless of how many light objects were dropped. Thus, grackles had to inhibit dropping any heavy objects to solve this problem, which made the task difficult. Despite the challenging apparatus, Refresco and Tequila likely would have further changed their preference to light objects if given more trials since their performance improved with the number of trials given, indicating that they were learning about the functional properties of the task.

Contrary to the only previous study comparing species with different brain sizes, which found a link between innovation frequency and brain size (Tebbich et al. 2010), I found no evidence to validate the link between innovation and behavioral flexibility or the link between behavioral flexibility and relative brain size when comparing non-innovative, average-brained great-tailed grackles with innovative, large-brained New Caledonian crows. Both species exhibited behavioral flexibility despite their differences in innovation frequency and relative brain size (Logan et al. 2014). My results are consistent with findings from the two studies on large-brained species that directly investigated the relationship between innovation frequency and behavioral flexibility, and behavioral flexibility and relative brain size (Auersperg et al. 2011; Manrique et al. 2013). I conclude that behavioral flexibility must be quantified directly in each species rather than using innovation as an indirect proxy since it is unclear what innovation frequency actually measures. Future research using proxies for behavioral flexibility at a broad taxonomic scale should choose a proxy other than innovation frequency and validate it across a number of species before relying on it. As the field stands now, it is unclear what the cross-species correlations between innovation frequency and other factors imply.

ACKNOWLEDGEMENTS

I am grateful to Luisa Bergeron, Christin Palmstrom, Linnea Palmstrom, and Michelle Gertsvolf for trapping and aviary assistance; Alexis Breen for helping to train Jugo; Steve Rothstein for scouting grackles and for use of the aviaries; Joe Jablonski and David Bothman for making the apparatuses; Jill Zachary and Kathy Frye at Santa Barbara City Parks and Recreation for use of the Andree Clark Bird Refuge; Estelle Sandhaus and Chris Briggs at the Santa Barbara Zoo for access to wild grackles; Karrie Black for managing purchasing and grants; Alex Thornton for logistical advice; Dieter Lukas for conceptual input and analysis feedback; Krista Fahy at the Santa Barbara Museum of Natural History for helping with permit applications; Kristine Johnson, Sarah Overington, Julie Morand-Ferron, and Neeltje Boogert for grackle advice and trap plans; Manny Garcia for veterinary and permit support; Mary Hunsicker and Bertrand Lemasson for assistance with making the trap; Margaret Tarampi, Eric Egenolf, Rebecca Schaefer, and Sam Franklin for brainstorming object designs; Andrew Iwaniuk for access to the brain size database; Will Hoppitt for GLMM effect size interpretation assistance; Irina Mikhalevich and Ljerka Ostojić for manuscript feedback; and my funders: the SAGE Center for the Study of the Mind at the University of California Santa Barbara and the National Geographic Waitt Grants Program.

REFERENCES

Akaike H (1981) Likelihood of a model and information criteria. J Econometrics 16: 3-14. doi: 10.1016/0304-4076(81)90071-3

Alma S, Bee De Speroni N (1992) Indices cerebrales y composicion cuantitativa encefalica en *Athene cunicularia* y *Tyto alba*. Facena 9:19-37.

Auersperg AMI, von Bayern AMP, Gajdon GK, Huber L, Kacelnik A (2011) Flexibility in problem solving and tool use of kea and New Caledonian crows in a multi access box paradigm. PLOS ONE 6:e20231. doi: 10.1371/journal.pone.0020231

Bates D, Maechler M, Bolker B (2011) lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-42. http://CRAN.R-project.org/package=lme4. Accessed 12 January 2015

Bee de Speroni N, Carezzano F (1995) Volumetric analysis of the visual, trigeminal and acoustic nuceli in four avian species (Rheidae, Spheniscidae, Tinamidae). Mar Ornithol 23: 11-15.

Bird CD, Emery NJ (2009) Rooks use stones to raise the water level to reach a floating worm. Curr Biol 19:1410-1414. doi: 10.1016/j.cub.2009.07.033

Boire D, Baron G (1994) Allometric comparison of brain and main brain subdivisions in birds. Journal of Brain Research 35:49-66

Buckner C (2013) A property cluster theory of cognition. Philosophical Psychology 28:307-336. doi: 10.1080/09515089.2013.843274

Burnham KP, Anderson DR (2002) Model selection and multimodel inference: a practical information- theoretic approach, 2nd edn. Springer, New York.

Carezzano FJ, Bee De Speroni N (1995) Composicion volumetrica encefalica e indices cerebrales en tres aves de ambiente acuatico (Ardeidae, Podicipedidae, Rallidae). Facena 11:75-83.

Cheke LG, Bird CD, Clayton NS (2011) Tool-use and instrumental learning in the Eurasian jay (*Garrulus glandarius*). Anim Cogn 14:441-455. doi: 10.1007/s10071-011-0379-4

Cheke LG, Loissel E, Clayton NS (2012) How do children solve Aesop's Fable? PLOS ONE 7:e40574. doi: 10.1371/journal.pone.0040574

Crile G, Quiring DP (1940) A record of the body weight and certain organ and gland weights of 3690 animals. Ohio J Sci 40:219-259.

Day LB, Westcott DA, Olster DH (2005) Evolution of bower complexity and cerebellum size in bowerbirds. Brain Behav Evol 66: 62-72.

Ducatez S, Clavel J, Lefebvre L (2015) Ecological generalism and behavioural innovation in birds: technical intelligence or the simple incorporation of new foods? J Anim Ecol 84:79-89. doi: 10.1111/1365-2656.12255

Ebinger P (1995) Domestication and plasticity of brain organization in mallards (*Anas platyrhynchos*). Brain Behav Evol 45: 286-300. doi: 10.1159/000113557

Fernandez P, Carezzano F, Bee De Speroni N (1997) Analisis cuantitativo encefalico e indices cerebrales en *Aratinga acuticaudata* y *Myiopsitta monachus* de Argentina (Aves: Psittacidae). Rev Chil Hist Nat 70:269-275.

Frahm HD, Rehkamper G (1998) Allometric comparison of the brain and brain structures in the white crested Polish chicken with uncrested domestic chicken breeds. Brain Behav Evol 52: 292-307. doi: 10.1159/000006574

Hadfield J (2010) MCMCglmm: Markov chain Monte Carlo methods for generalised linear mixed models. http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=E395FF7367BF580D916E465F55302C60?doi=10.1.1.160.5098&rep=rep1&type=pdf Accessed 7 May 2015

Hadfield J (2014) MCMCglmm course notes. http://cran.r-project.org/web/packages/MCMCglmm/vignettes/CourseNotes.pdf. Accessed 7 May 2015

Hadfield JD, Nakagawa S (2010) General quantitative genetic methods for comparative biology: Phylogenies, taxonomies, meta-analysis and multi-trait models for continuous and categorical characters. J Evol Biol 23:494–508.

Healy SD, Rowe C (2007) A critique of comparative studies of brain size. Proc R Soc B 274:453-464. doi: 10.1098/rspb.2006.3748

Hunt GR (1996) Manufacture and use of hook-tools by New Caledonian crows. Nature 379:249-251. doi: 10.1038/379249a0

Hunt GR (2008) Introduced *Lantana camara* used as tools by New Caledonian crows (*Corvus moneduloides*). N Z J Zool 35:115-118.

Hunt GR, Gray RD (2004) Direct observations of pandanus-tool manufacture and use by a New Caledonian crow (*Corvus moneduloides*). Anim Cogn 7:114-120. doi: 10.1007/s10071-003-0200-0

Iwaniuk AN (2004) Brood parasitism and brain size in cuckoos: A cautionary tale on the use of modern comparative methods. Int J Comp Psychol 17: 17-33.

Iwaniuk AN, Dean KM, Nelson JE (2005) Interspecific allometry of the brain and brain regions in parrots (Psittaciformes): comparisons with other birds and primates. Brain Behav Evol 65:40-59. doi: 10.1159/000081110

Iwaniuk AN, Hurd PL (2005) The evolution of cerebrotypes in birds. Brain Behav Evol 65:215-230. doi: 10.1159/000084313

Iwaniuk AN, Nelson JE (2001) A comparative analysis of relative brain size in waterfowl (Anseriformes). Brain Behav Evol 57:87-97. doi: 10.1159/000047228

Iwaniuk AN, Nelson JE (2003) Developmental differences are correlated with relative brain size in birds: a comparative analysis. Can J Zool 81: 1913-1928.

Iwaniuk AN, Nelson JE, James HF, Olson SL (2004) A comparative test of the correlated evolution of flightlessness and relative brain size in birds. J Zool Lond 263: 317-327. doi: 10.1017/S0952836904005308

Iwaniuk AN, Wylie DRW (2006) The evolution of stereopsis and the Wulst in caprimulgiform birds: a comparative analysis. J Comp Physiol A 192: 1313-1326. doi: 10.1007/s00359-006-0161-2

Jelbert SA, Taylor AH, Cheke LG, Clayton NS, Gray, RD (2014) Using the Aesop's fable paradigm to investigate causal understanding of water displacement by New Caledonian crows. PLOS ONE 9:e92895. doi: 10.1371/journal.pone.0092895

Layard EL, Layard ELC (1882) Notes on the avifauna of New Caledonia. Ibis 6:520-522.

Lefebvre L, Whittle P, Lascaris E, Finkelstein A (1997) Feeding innovations and forebrain size in birds. Anim Behav 53:549-560. doi: 10.1006/anbe.1996.0330

Lefebvre L, Reader SM, Sol D (2004) Brains, innovation and evolution in birds and primates. Brain Behav Evol 63:233-246. doi: 10.1159/000076784

Lefebvre L, Reader SM, Sol D (2013) Innovating innovation rate and its relationship with brains, ecology and general intelligence. Brain Behav Evol 81:143-145. doi: 10.1159/000348485

Le Goupils, M., 1928. Dans la Brousse Calédonienne: Souvenirs d'un Ancien Planteur, 1898-1904. Perrin, Paris.

Logan CJ, Jelbert SA, Breen AJ, Gray RD, Taylor AH (2014) Modifications to the Aesop's Fable paradigm change performances in New Caledonian crows. PLOS ONE 9:e103049. doi:10.1371/journal.pone.0103049

Logan C (2015) Great-tailed grackle water tube experiments, Santa Barbara, CA USA 2014-2015. KNB Data Repository. doi:

Manrique HM, Völter CJ, Call J (2013) Repeated innovation in great apes. Anim Behav 85:195-202. doi: 10.1016/j.anbehav.2012.10.026

Mlikovsky J (1989a) Brain Size in Birds: 1. Tinamiformes Through Ciconiformes, Vestrik Ceskoslovenska Spolecnosti Zoologicka 53:33-47.

Mlikovsky J (1989b) Brain Size in Birds: 2. Falconiformes through Gaviiformes. Vestrik Ceskoslovenska Spolecnosti Zoologicka 53:200-213.

Mlikovsky J (1989c) Brain Size in Birds: 3. Columbiformes Through Piciformes. Vestrik Ceskoslovenska Spolecnosti Zoologicka 53 252-273.

Mlikovsky J (1990) Brain Size in Birds: 4. Passeriformes. Acta Soc Zool Bohemoslov 54:27-101.

Møller AP, Erritzoe J, Garamszegi LZ (2004) Covariation between brain size and immunity in birds: implications for brain size evolution. J Evol Biol 18:223-237. doi: 10.1111/j.1420-9101.2004.00805.x

Nicolakakis N, Lefebvre L (2000) Forebrain size and innovation rate in European birds: feeding, nesting and confounding variables. Behav 137:1415-1429. doi: 10.1163/156853900502646

Overington SE, Morand-Ferron J, Boogert NJ, Lefebvre L (2009) Technical innovations drive the relationship between innovativeness and residual brain size in birds. Anim Behav 78:1001-1010. doi: 10.1016/j.anbehav.2009.06.033

Overington SE, Cauchard L, Cote KA, Lefebvre L (2011) Innovative foraging behavior in birds: What characterizes an innovator? Behav Process 87:274-285. doi: 10.1016/j.beproc.2011.06.002

Pistone E, Carezzano F, Bee De Speroni N (2002) Tamano relativo encefalico e indices cerebrales en *Vanellus c. chilensis* (Aves: Charadriidae). Rev Chil Hist Nat 75: 595-602. doi: 10.4067/S0716-078X2002000300010

Pyle P (2001) Identification guide to North American birds part 1. Sheridan Books, Inc., Ann Arbor, Michigan.

R Core Team (2014) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org. Accessed 3 Jan 2015.

Reader SM, Laland KN (2002) Social intelligence, innovation, and enhanced brain size in primates. Proceedings of the National Academy of Sciences 99:4436-4441. doi: 10.1073/pnas.062041299

Rehkämper G, Frahm HD, Cnotka J (2008) Mosaic evolution and adaptive brain component alteration under domestication seen on the background of evolutionary theory. Brain Behav Evol 71: 115-126. doi: 10.1159/000111458

Rehkämper G, Frahm HD, Zilles K (1991) Quantitative development of brain and brain structures in birds (Galliformes and Passeriformes) compared to that in mammals (Insectivores and Primates). Brain Behav Evol 37:125-143. doi: 10.1159/000316088

Rehkämper G, Kart E, Frahm HD, Werner CW (2003) Discontinuous variability of brain composition among domestic chicken breeds. Brain Behav Evol 61: 59-69.

Taylor AH, Elliffe DM, Hunt GR, Emery NJ, Clayton NS, Gray RD (2011) New Caledonian crows learn the functional properties of novel tool types. PLOS ONE 6:e26887. doi:10.1371/journal.pone.0026887

Tebbich S, Sterelny K, Teschke I (2010) The tale of the finch: adaptive radiation and behavioural flexibility. Philos Trans R Soc B 365:1099-1109. doi: 10.1098/rstb.2009.0291

Timmermans S, Lefebvre L, Boire D, Basu P (2000) Relative size of the hyperstriatum ventrale is the best predictor of feeding innovation rate in birds. Brain Behav Evol 56:196-203. doi: 10.1159/000047204