

1 **Innovation frequency does not indicate behavioral flexibility in great-tailed grackles**

2 Corina J. Logan*

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

4 93103 USA

5 *Current address: Department of Zoology, University of Cambridge, Cambridge CB2 3EJ United

6 Kingdom

7 E-mail: cl417@cam.ac.uk, phone: +44 7868 357 294, fax: +44 1223 336 676

8
9 **ABSTRACT**

10 Many cross-species studies attest that innovation frequency (novel food types eaten and foraging
11 techniques used) is a measure of behavioral flexibility and show that it positively correlates with
12 relative brain size (corrected for body size). However, mixed results from the three studies that
13 directly test the relationship between innovation frequency and behavioral flexibility and behavioral
14 flexibility and brain size question both assumptions. I investigated behavioral flexibility in non-
15 innovative great-tailed grackles that have an average sized brain, and compared their test
16 performance with innovative, large-brained New Caledonian crows. Contrary to the prediction,
17 grackles perform similarly to crows in experiments using clear tubes partially filled with water and
18 containing a floating food reward, where objects must be dropped into the tube to raise the water
19 level, bringing the food within reach. Similarly to crows, 4 out of 6 grackles preferred to drop the
20 more functional heavy (rather than light) objects, and 2 changed their preference in a follow up
21 experiment where the heavy objects were no longer functional, thus exhibiting behavioral
22 flexibility. These results challenge the assumption that innovation frequency indicates behavioral
23 flexibility since a non-innovative bird demonstrated behavioral flexibility at a level similar to that in
24 innovative crows, and they challenge the assumption that only large brains are capable of
25 behavioral flexibility because a bird with an average brain size solved problems similarly to large-
26 brained crows.

27

28 **KEYWORDS**

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

31

32 **INTRODUCTION**

33 Behavioral flexibility, currently defined as the ability to adapt behavior to changing contexts, is
34 considered the keystone of complex cognition (Buckner 2013). Measuring behavioral flexibility
35 directly in each species is time intensive. Thus, comparative biologists seek behaviors that can serve
36 as indicators of behavioral flexibility, therefore allowing cross-species comparisons of cognition
37 with behavior, ecology, and life history (Lefebvre et al. 1997; see Healy and Rowe 2007 for a
38 review). One widely used indicator of behavioral flexibility is the frequency of innovations, where
39 innovations are based on reports of novel food types eaten and foraging techniques used (Lefebvre
40 et al. 1997, 2004, 2013; Timmermans et al. 2000; Nicolakakis and Lefebvre 2000; Reader and
41 Laland 2002). This operational definition of behavioral flexibility relies on two main assumptions:
42 1) innovativeness indicates complex cognition through behavioral flexibility and 2) innovativeness
43 actually measures behavioral flexibility (e.g., Lefebvre et al. 1997, 2002; Timmermans et al. 2000).
44 The first assumption presumes that behavioral flexibility can serve as evidence of complex
45 cognition because those species with more information processing capacity (as indicated by relative
46 brain size [corrected for body size]) should be able to adapt their behavior more flexibly to
47 changing circumstances (e.g., Jerison 1985). There is evidence that innovation frequencies correlate
48 with relative brain size across species (Lefebvre et al. 1997, 2004, 2013; Timmermans et al. 2000;
49 Nicolakakis and Lefebvre 2000; Reader and Laland 2002), however it is a circular argument to use
50 behavioral flexibility as an indicator of complex cognition when complex cognition itself is defined
51 as flexible behavior. This circular argument needs validation from external factors (Healy & Rowe
52 2007). Without this validation, progress can only be made on the second assumption. Here, I review

53 the evidence provided by direct tests of behavioral flexibility in species that vary in their innovation
54 frequencies and relative brain sizes, and I test the assumption that innovation frequency is a proxy
55 for behavioral flexibility by comparing two bird species that differ in both respects.

56 It is important to clarify the distinction between two categories of innovation that are often
57 referred to: innovation frequency is the number of innovations per species, while innovativeness
58 refers to an individual's propensity to innovate. The literature on innovativeness directly measures
59 this variable at the individual level and relates it to other quantified behaviors (e.g., problem
60 solving) at taxonomic scales appropriate given the scale at which experiments are conducted (see
61 Griffin & Guez 2015 for a review). In this paper, I solely refer to innovation frequencies per species
62 because it is this literature that makes the assumptions previously stated, as well as suffering from
63 other drawbacks (e.g., anecdotal reports, biased toward the visibility of the species, Healy & Rowe
64 2007).

65 Results from experiments testing behavioral flexibility show mixed evidence from cross-
66 species studies about how behavioral flexibility relates to innovation frequency and relative brain
67 size. There is some evidence that behavioral flexibility correlates with innovation frequency, but not
68 relative brain size: Innovative, smaller-brained Galapagos finches reversed a previously learned
69 color preference faster than less innovative, relatively larger-brained New World jays (Tebbich et
70 al. 2010). Other evidence shows that behavioral flexibility correlates with relative brain size, but not
71 innovation frequency: Keas and New Caledonian crows performed similarly on a multi-access box
72 — where the successful solution of an option resulted in its closure, thereby forcing the individual
73 to innovate another solution on the same box — even though the crows are reported to have more
74 innovations than the keas (Auersperg et al. 2011). Evidence from primates shows that behaviorally
75 flexible problem solving does not correlate with innovation frequency or relative brain size:
76 Chimpanzees, bonobos, and gorillas performed better on a multi-access box than orang-utans, and
77 all species (except for all but one orang-utan) quickly moved on to other techniques for accessing
78 the food when the current method stopped working (Manrique et al. 2013). All of these primate

79 species are innovative and have relatively large brains, with gorillas having the smallest of this
80 group of species (Isler et al. 2008). The multi-access box experiment measures behavioral flexibility
81 in successful individuals by requiring them to adapt their behavior to changing circumstances.
82 Therefore, so far, direct evidence indicates that innovation frequencies per species may not measure
83 behavioral flexibility and that behavioral flexibility does not correlate with relative brain size. The
84 one study that found evidence supporting the link between innovation frequency and behavioral
85 flexibility comes from the only study to compare species with different brain sizes (Tebbich et al.
86 2010). The other two studies examined relatively large brained species and found a consistent lack
87 of support regarding the link between innovation frequency and behavioral flexibility. Therefore, it
88 is unclear how innovation frequency relates to brain size when examining behavioral flexibility
89 directly.

90 The aim of my study was to test the relationship between innovation frequency and
91 behavioral flexibility, and behavioral flexibility and relative brain size in species with differing
92 brain sizes by directly measuring behavioral flexibility in great-tailed grackles (*Quiscalus*
93 *mexicanus*) and comparing their performance with previously tested New Caledonian crows
94 (*Corvus moneduloides*; Logan et al. 2014). Grackles are not innovative (n=1 innovation; Ducatez et
95 al. 2014) and have an average relative brain size, while New Caledonian crows are innovative
96 (n=12 innovations; literature reviewed by Logan: Layard & Layard 1882, Le Goupils 1928, Hunt
97 1996, 2000, 2008, Hunt & Gray 2002 and 2004, Rutz et al. 2007, Troscianko et al. 2008) and have
98 relatively large brains. Therefore, the prediction is that the less innovative and smaller-brained
99 grackles will show less behavioral flexibility than crows. I obtained relative brain size data from a
100 database including 2131 bird species because this was the only database that included both great-
101 tailed grackles and New Caledonian crows, which is essential for determining their brain sizes
102 relative to each other, and for confirming whether the grackle brain is of average size for a bird
103 (Alma and Bee de Speroni 1992; Bee de Speroni and Carezzano 1995; Boire and Baron 1994;
104 Carezzano et al. 1995; Crile and Quiring 1940; Day et al. 2005; Ebinger 1995; Ebinger and Lohmer

105 1984, 1987; Ebinger and Rohrs 1995; Fernandez et al. 1997; Frahm and Rehkamper 1998, 2004;
106 Iwaniuk and Nelson 2003; Iwaniuk 2004; Iwaniuk et al. 2004, 2005; Iwaniuk and Wylie 2006;
107 Iwaniuk unpublished data [380 species]; Milkovsky 1989a,b,c, 1990; Møller et al. 2004; Pistone et
108 al. 2002; Rehkamper et al. 1991, 2003, 2008).

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

120 In this study, I modified the water tube experiments to investigate behavioral flexibility. I
121 tested behavioral flexibility, the ability to change preferences when the context changes (Buckner
122 2013), by presenting the grackles first with the heavy vs. light experiment and then with a follow up
123 experiment in which the heavy objects were no longer functional. In this follow up experiment
124 (heavy vs. light magic), heavy objects stuck to a magnet placed inside the tube above the water
125 level, leaving the light objects as the functional option because they could fall past the magnet and
126 into the water. If grackles preferred heavy objects or had no preference in the heavy vs. light
127 experiment and then changed their preference in the heavy vs. light magic experiment to preferring
128 neither object or light objects, this would indicate that their preferences are sensitive to changing
129 contexts. New Caledonian crows exhibited behavioral flexibility using the water tube tests when
130 they discriminated between two tubes of different volumes (Logan et al. 2014). In the first

131 experiment, crows preferred to drop objects into a narrow (functional) rather than a wide (non-
132 functional) tube when water levels were equal in both tubes. In a follow up experiment where the
133 narrow tube was no longer functional because the water level was too low, crows changed their
134 preference to dropping objects into the functional wide tube. I carried out these same experiments
135 with the grackles to compare their flexibility with that in New Caledonian crows.

136 To summarize, behavioral flexibility would be shown if the grackles that preferred heavy in
137 heavy vs. light changed their preference to no preference or to preferring light objects in heavy vs.
138 light magic experiment, and if those grackles that preferred the narrow tube in narrow vs. wide with
139 equal water levels experiment changed their preference to the wide tube in narrow vs. wide with
140 unequal water levels experiment. This paradigm is similar to reversal learning experiments, which
141 are considered tests of behavioral flexibility (e.g., Bond et al. 2007, Tebbich et al. 2010,
142 Ghahremani et al. 2010, Buckner 2013). The crows were not given the heavy vs. light magic
143 experiment because it had not been designed yet, therefore grackle and crow behavioral flexibility
144 could be directly compared using the wide vs. narrow equal and unequal water level experiments.
145 Behavioral flexibility in these two species could be more generally compared in terms of their
146 ability to change preferences when circumstances change regardless of which experiments they
147 demonstrate flexibility in.

148

149 **METHODS**

150 **Ethics**

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

156

157 **Subjects and Study Site**

158 Eight wild adult great-tailed grackles (4 females and 4 males) were caught using a walk-in baited
159 trap measuring 0.61m high by 0.61m wide by 1.22m long (design from Overington et al. 2011).
160 Birds were caught (and tested) in two batches: batch one at the Andree Clark Bird Refuge (4 birds
161 in September 2014, released in December) and batch two at East Beach Park (4 birds in January
162 2015, released in March) in Santa Barbara, California. They were housed individually in aviaries
163 measuring 183cm high by 119cm wide by 236cm long at the University of California Santa Barbara
164 for 2-3 months while participating in the experiments in this study. Grackles were given water *ad*
165 *libitum* and unrestricted amounts of food (Mazuri Small Bird Food, bread, and peanuts) for at least
166 20 hrs per day, with their main diet being removed for up to 4 hrs on testing days while they
167 participated in experiments. Grackles were aged by plumage and eye color and sexed by plumage
168 and weight following Pyle (2001). Biometrics, blood, and feathers were collected at the beginning
169 and end of their time in the aviary. Their weights were measured at least once per month, first at the
170 time of trapping using a balancing scale, and subsequently by placing a kitchen scale covered with
171 food in their aviary and recording their weight when they jumped onto the scale to eat.

172

173 **Color Learning to Prevent Side Bias**

174 To help break potential side biases during the wide vs. narrow water tube experiment, I first had
175 grackles learn a simple association between food and color, which forced them to pay attention to
176 color rather than spatial location (see Logan et al. 2014). They were given a silver and a gold tube
177 with food always hidden in the gold tube. One silver and one gold tube were placed at opposite ends
178 of a table with the tube openings facing the side walls so the bird could not see which tube
179 contained the food. Tubes were pseudorandomized for side and the left tube was always placed on
180 the table first, followed by the right to avoid behavioral cueing. Pseudorandomization involved
181 alternating sides for the first two trials in a 10-trial set and then never having one tube on the same
182 side for more than two trials in a row, while avoiding a pattern that would allow the bird to follow a

183 rule to solve the task rather than learning which color indicated the food. Each trial consisted of
184 placing the tubes on the table and then the bird had the opportunity to choose one tube by looking
185 into it (and eating from it if it chose the gold tube). Once the bird chose, the trial ended by
186 interrupting the bird and removing the tubes. A bird passed this test if it made at least 17 correct
187 choices out of the most recent 20 trials. Proficiency with this test then served as a useful tool for
188 later water tube experiments involving two tubes: if a grackle developed a side bias, the water tube
189 experiment was paused and silver/gold tests were conducted until the bird attended to color rather
190 than location (side).

191

192 **Spontaneous Stone Dropping**

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

202

203 **Stone Dropping Training**

204 Those birds that did not spontaneously drop stones down the tube on the stone dropping training
205 apparatus were trained to push or drop stones down tubes using this apparatus. Birds were given
206 two stones and went from accidentally dropping stones down the tube as they pulled at food under
207 the stones, which were balanced on the edge of the tube opening, to pushing or dropping stones into
208 the tube from anywhere near the apparatus. Once the bird proficiently pushed or dropped stones into

209 the apparatus 30 times, they moved onto the reachable distance on a water tube. Stone
210 pushing/dropping proficiency was defined as consistently directing the stone to tube opening from
211 anywhere on the ramp on the top of the apparatus. Not all motions had to be in the direction of the
212 tube opening because some grackles preferred to move the stone to a particular location on the ramp
213 (which may initially be in the opposite direction from the tube) and push or drop it in from there or
214 push the stone in shorter, angular strokes. It was permissible for a bird to throw one of the stones off
215 the side of the apparatus (which occurred sporadically throughout all of their experiences with stone
216 pushing/dropping) as long as they proficiently put the other stone in the tube.



217

218 Figure 1. Batido participates in stone dropping training.

219

220

221 **Reachable Distance**

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

230

231 **Water Tube Proficiency Assessment**

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

238

239 **Experimental Set Up**

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

246

247 **Experiment 1: Heavy vs. Light**

248 A water tube was presented with 4 heavy (steel rod wrapped in fimo clay, weight=10g, each
249 displaces 2-3mm of water) and 4 light (plastic tube partially filled with fimo clay, weight=2g, each
250 displaces 1-1.5mm of water) objects placed in pseudorandomized (as explained for color learning)
251 pairs near the top of the tube (both objects were 21-24mm long and 8mm in diameter; Figure 2).
252 Heavy objects displaced 0.5-2mm more water than light objects, thus making them more functional
253 than the light objects, but importantly, both objects were functional. Each bird had three
254 opportunities to interact with the objects before the experiment began: one heavy and one light
255 object was placed on the table (pseudorandomized for side) with food underneath and on top of
256 each object. The object that was first touched was recorded and a trial continued until the bird
257 interacted with both objects. If one object was preferred (as indicated by approaching it first more
258 than once), then more food was placed on the other object to try to eliminate any object preference
259 before the experiment began.



260
261 Figure 2. The Heavy vs. Light experimental set up.

262

263

264

265 **Experiment 2: Heavy vs. Light Magic**

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



271

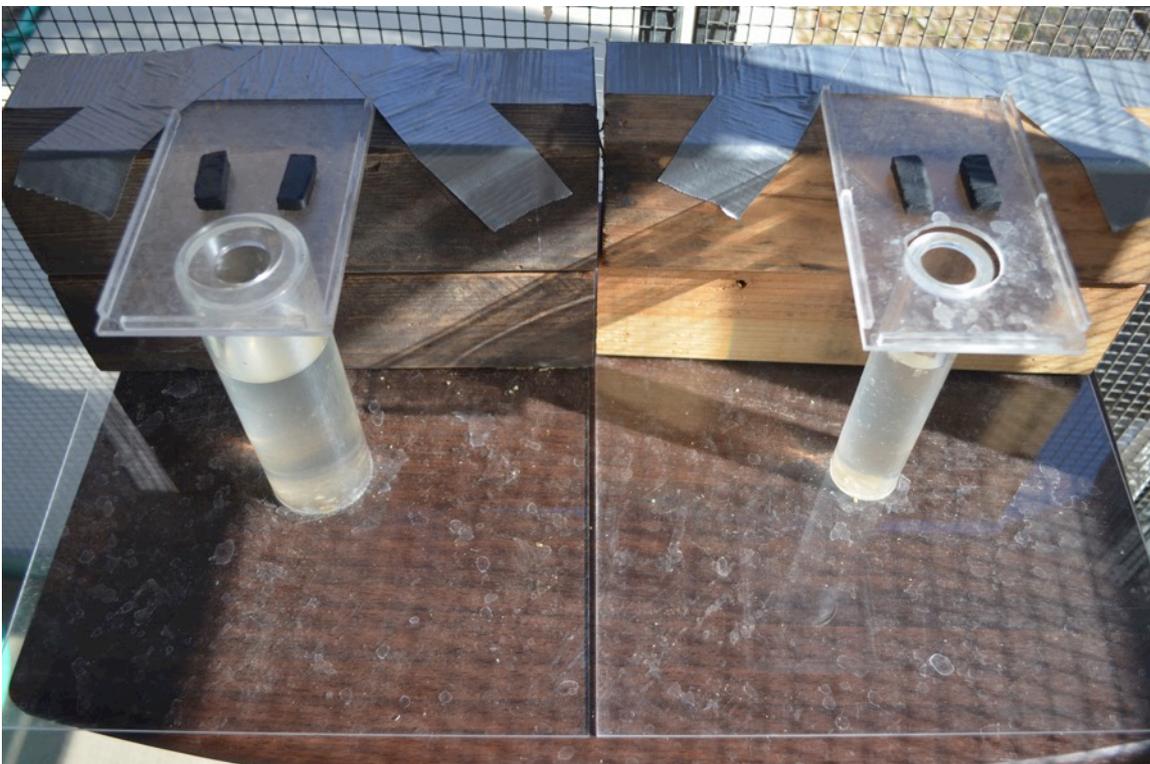
272 Figure 3. The Heavy vs. Light Magic experimental set up, which includes magnets stuck to the tube
273 above the water level.

274

275 **Experiment 3: Narrow vs. Wide Equal Water Levels**

276 To determine whether birds understand volume differences, a wide and narrow tube with equal
277 water levels were presented with four objects made out of fimo clay (30x10x5mm, 3-4g, each
278 object displaced 1-2mm in wide tube and 5-6mm in narrow; Logan et al. 2014; Figure 4). Two
279 objects were placed near the narrow tube opening and two objects near the wide tube opening. The
280 objects were only functional if dropped into the narrow tube because the water levels were set such
281 that dropping all of the objects into the wide tube would not bring the floating food within reach.
282 However, dropping 1-2 objects into the narrow tube would raise the water level enough to reach the

283 food. Both tubes were 170mm tall with 3mm thick lids that constricted the opening to 25mm in
284 diameter to equalize the bird's access to the inside of each tube, and super glued to a clear acrylic
285 base (300x300x3mm). The wide tube (outer diameter=57mm, inner diameter=48mm,
286 volume=307,625mm³) was roughly equally larger than the standard water tube (dimensions above,
287 volume=192,800mm³) as the narrow tube was smaller (outer diameter=38mm, inner
288 diameter=25mm, volume=83,449mm³). The position of the tubes was pseudorandomized for side to
289 ensure that tube choices were not based on a side bias. Before the experiment began, each bird had
290 three opportunities to interact with the object as in Experiment 1.



291
292 Figure 4. The Narrow vs. Wide with equal water levels experimental set up.

293

294 **Experiment 4: Narrow vs. Wide Unequal Water Levels**

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

300

301 **Experimenters**

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

304

305 **Statistical Analyses**

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

311 Generalized linear mixed models (GLMMs) were used to determine whether birds preferred
312 particular objects or tubes (response variable: correct or incorrect choice) in an experiment and
313 whether the trial number or bird influenced choices (explanatory variables: experiment, trial
314 number, bird), and to control for the non-independence of multiple choices per trial (random factor:
315 choice number). I set the prior to fix the variance component to one ($\text{fix}=1$) because the
316 measurement error variance was known, as is standard when choices are binary (Hadfield 2010). I
317 ensured that the Markov chain for this test model converged by manipulating the number of
318 iterations ($\text{nitt}=150000$), the number of iterations that must occur before samples are stored
319 ($\text{burnin}=30000$), and the intervals the Markov chain stores ($\text{thin}=300$) until successive samples were
320 independent (autocorr function, MCMCglmm package: Hadfield 2010) and there were no trends
321 when visually inspecting the time series of the Markov chain (function: `plot(testmodel$Sol)`;
322 Hadfield 2014). I compared this test model to a null model where I removed all explanatory factors
323 and set it to 1. I determined whether the test model was likely given the data, relative to the null
324 model by using Akaike weights (range: 0-1, all model weights sum to 1; Akaike 1981; Weights
325 function, MuMIn package: Bates et al. 2011). The Akaike weight indicates the “relative likelihood

326 of the model given the data” (Burnham and Anderson 2002, p. xxiii) and models with Akaike
327 weights greater than 0.9 are considered reliable models because they are highly likely given the data
328 (Burnham and Anderson 2002). The test model was highly likely given the data (Akaike
329 weight=0.99) and the null model was not (Akaike weight=0.0009). To investigate potential effects
330 of season or order of testing, I carried out a GLMM to investigate whether the batch to which the
331 bird belonged (explanatory variable: batch=1 or 2) influenced their test performance (response
332 variable: correct or incorrect choice) while controlling for the non-independence of multiple choices
333 per trial (random factor: choice number). The null model was highly likely given the data (Akaike
334 weight=0.94), while the batch model was not (Akaike weight=0.06), indicating that batch did not
335 influence test performance. GLMMs were carried out in R v3.1.2 (R Core Team 2014) using the
336 MCMCglmm function (MCMCglmm package) with a binomial distribution (called categorical in
337 MCMCglmm) and logit link.

338

339 **Data Availability**

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

342

343 **RESULTS**

344 **Spontaneous Stone Dropping**

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

347

348 **Stone Dropping Training**

349 Most grackles learned to push stones into a tube on the platform apparatus in 165-392 trials (Table
350 1), however Michelada was scared of the stone falling down the tube and did not habituate to this
351 event and Jugo learned too slowly to become proficient by the time he needed to be released,

352 therefore they were excluded from the stone dropping experiments. The training procedure was
353 modified from Logan et al. (2014) to allow stone pushing from a clear cast acrylic ramp placed on
354 top of the tube rather than stone dropping by picking up the stone from the table and putting it into
355 the tube (without a ramp; see Figures 1-6). The modification was necessary because grackles seem
356 to form associations between the stones and the top of the tube, the stones and the table where the
357 food comes out, and the stones falling only in one direction: down. When I placed the stones below
358 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
359 the tube, the grackles took the stones and dropped them off the side of the apparatus or table, often
360 placing them on the table and then looking at where the platform should have fallen open, awaiting
361 the food. Placing the ramp on the water tubes for the experiments was implemented to mitigate this
362 limitation. Once this change was made, it was no longer necessary to train the grackles to pick up
363 and drop the stones because pushing them into the tube sufficed and required less training.

364

365 **Water Tube Proficiency Assessment**

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

375

376 **Experiment 1: Heavy vs. Light**

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

390

391 **Experiment 2: Heavy vs. Light Magic**

392 Tequila and Refresco changed from preferring heavy objects in Experiment 1 to having no
393 preference in this experiment, while Batido continued to prefer the non-functional heavy objects
394 (see Table 1 for binomial test results and Table 2 for GLMM results). Margarita continued to prefer
395 heavy items and Cerveza went from having no preference to preferring the non-functional heavy
396 items, likely due to their interest in the magnet (Table 1). The magnet seemed to attract their
397 interest, thus continuing or increasing their preference for heavy objects (Supplementary Material
398 Video 2 [also available at: <https://youtu.be/TrKWEch1Y5M>]). Tequila gave up after 17 trials,
399 refusing to drop either type of object into the tube, indicating he may have inhibited his choice of
400 heavy. Tequila and Refresco's performance improved with trial number, indicating that they learned
401 through trial and error about which object was functional (Table 2). The other grackles
402 performances decreased with increasing trial number, indicating that they did not learn about which

403 object was functional (Table 2). Even though Tequila and Refresco did not learn to prefer light in
 404 the amount of trials given, they did exhibit flexibility in that they changed their preferences from
 405 heavy in the previous experiment to having no preference in this experiment. Indeed, Refresco
 406 would likely have shown a preference for light objects if given more trials since all choices in his
 407 last five trials were light objects (Supplementary Material Figure S1.2).

408

409 Table 1. Performance per bird per experiment: the number of stone dropping training trials needed
 410 to reach proficiency, and p-values from Bonferroni-Holm corrected (within experiment) binomial
 411 tests for each experiment (- = was not given this experiment). Note: Tequila was the first bird tested
 412 and I did not realize until after I trained him to pick up and drop the stones into the tube that I
 413 wanted to only train the other birds to push the stones into the tube to save training time. Therefore,
 414 the trial numbers for the other birds refer to proficiency to push objects into the tube, not pick up
 415 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	-	-	-	-

416

417 Experiment 3: Narrow vs. Wide Equal Water Levels

418 All three grackles that participated in this experiment displayed no preference for dropping objects
 419 into the functional narrow tube or the non-functional wide tube (see Table 1 for binomial test results
 420 and Table 2 for GLMM results, Supplementary Material Video 3 [also available at:
 421 <https://youtu.be/25Dj3vnSz5M>]). None of the grackles' performances improved with trial number,

422 indicating that they did not learn to distinguish which one was functional (Table 2). Batido appeared
423 to rely on the strategy of dropping all objects into both tubes regardless of which tube he received a
424 reward from, although in trial 12, he picked up the objects from the wide tube area and dropped
425 them into the narrow tube even though he was only trained to push stones, not drop them
426 (Supplementary Material Figure S1.3). Since no grackle passed this experiment, they were not
427 given experiment 4 (Narrow vs. Wide with unequal water levels), which would have investigated
428 their behavioral flexibility in this context.

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

441

442 **First Choices on First Trials**

443 All six grackles chose the more functional heavy objects as their first choice in their first trial in
444 Heavy vs. Light, which indicates that they preferred the heavy objects from the very beginning of
445 the experiment (Figure S1.1). Five out of six grackles chose the non-functional heavy objects in
446 Heavy vs. Light Magic (Figure S1.2), which is not surprising given that they had learned to prefer
447 heavy objects in the previous experiment and had likely never interacted with a magnet before,

448 therefore they should have had no reason to have a prior understanding of how the Magic
 449 experiment worked. Two out of three grackles chose the functional narrow tube in Narrow vs. Wide
 450 with equal water levels, indicating no initial preference for a particular tube (Figure S1.3).

451

452 Table 2. Examining the influence of experiment, trial, and bird on test success (Test Performance)
 453 and whether success increased with trial number (Learning Effects), thus indicating a learning
 454 effect. GLMM: Choices Correct ~ Experiment*Trial*Bird, random = ~Choice Number. CI=credible
 455 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	-	-	-
<i>Heavy vs. Light</i>						
<i>Batido</i>	<i>1.27</i>	<i>0.08</i>	<i>2.65</i>	-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
<i>Heavy vs. Light Magic</i>						
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
<i>Narrow vs. Wide</i>						
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

456

457

458 **Did choice number influence the results?**

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

464

465 **DISCUSSION**

466 Despite their average brain size and lack of innovations, great-tailed grackles performed similarly to
467 innovative and large-brained New Caledonian crows (Logan et al. 2014) on the Heavy vs. Light
468 experiment (see New Caledonian crow data in Table 3 for comparison), and grackles exhibited
469 behavioral flexibility by changing their preferences in the Heavy vs. Light Magic experiment.
470 Grackles and crows exhibited behavioral flexibility in different two-step experiments making it
471 difficult to directly compare how similar their behavioral flexibility is: crows were not given Heavy
472 vs. Light Magic because it was not invented yet and grackles were not able to be given Narrow vs.
473 Wide with unequal water level experiment because no grackle passed the equal water level
474 precursor. However, the fact that both species exhibited behavioral flexibility (Tables 1 and 3) using
475 the water tube paradigm allows for a more general comparison of behavioral flexibility as it relates
476 to a species' innovation frequency and brain size.

477 Both grackles and crows preferred to drop the functional heavy objects into the water tube
478 rather than the less functional (for grackles) or non-functional (for crows) light objects. All grackles
479 were successful at obtaining the food, and 4 out of 6 grackles preferred to drop the more functional
480 heavy objects. Similar to the crows (Table 3), Tequila, Batido, and Refresco preferred heavy objects
481 significantly more than light objects and did not show a learning effect across the 20 trials in this
482 experiment, indicating that they relied on prior information about the world to solve this task.

483 Table 3. Summary of New Caledonian crow performances based on p-values from Bonferroni-
 484 Holm corrected (within experiment) binomial tests in previous water tube experiments (Logan et al.
 485 2014: birds 1-6, Jelbert et al. 2014: birds 7-12, Taylor et al. 2011: birds 13-16). J = juvenile, A =
 486 adult, ^ = this experiment was called Sinking vs. Floating in Jelbert et al. (2014), * = data not
 487 directly comparable because these crows underwent a different stone dropping training regime,
 488 requiring 1-116 trials of stick pushing, bill pushing, or stone nudging (only the latter training was
 489 used on the grackles) on the stone dropping training apparatus (Logan unpublished data), ~ = Jelbert
 490 et al. (2014) gave the birds enough objects to be successful in either tube, - = was not given this
 491 experiment. Note that the crows did not have the heavy vs. light magic experiment because it was
 492 not invented yet, and the stone drop training trials for birds 13-16 were the number of accidental
 493 stone drops required before the bird began dropping stones.

	Bird	Sex	Age	Stone drop training trials	Heavy vs. Light[^]	Wide vs. Narrow Equal	Wide vs. Narrow Unequal
1	Q	M	J		0.004 heavy	0.02 narrow	0.02 wide
2	007	F	J		0.004 heavy	1.00	-
3	Kitty	F	J	1-116*	<0.001 heavy	0.03 narrow	0.30
4	Lady	M	A		<0.001 heavy	0.02 narrow	<0.001 wide
5	Buster	F	A		<0.001 heavy	<0.001 narrow	<0.001 wide
6	Damien	M	J		<0.001 heavy	1.00	-
7	R	M	A	*	<0.001 heavy	<0.001 wide~	<0.001 wide
8	W	F	A	*	<0.001 heavy	<0.001 wide~	<0.001 wide
9	Y	M	A	*	<0.001 heavy	<0.001 wide~	<0.001 wide
10	O	F	J	*	<0.001 heavy	<0.001 wide~	<0.001 wide
11	RB	F	J	*	<0.001 heavy	<0.001 wide~	<0.001 wide
12	WG	F	J	*	<0.001 heavy	<0.001 wide~	<0.001 wide
13	Caesar	M	A		<0.001 heavy	-	-
14	Laura	F	A	12.25 mean ±6.9 s.e.m.	<0.001 heavy	-	-
15	Mimic	M	J		<0.001 heavy	-	-
16	Pepe	M	J		<0.001 heavy	-	-

494 Behavioral flexibility was exhibited by grackles insofar as they changed their preferences
 495 when the task changed. When the heavy objects in the Heavy vs. Light Magic experiment were no
 496

497 longer functional because they stuck to a magnet, two grackles changed from having preferred
498 heavy objects when they were functional in the previous experiment to having no object preference
499 in the Magic experiment. This demonstrates attention to the functional properties of objects in
500 changing circumstances. No grackle completely switched their preference to the light objects, which
501 may have been made difficult by the design of the apparatus: if one heavy item was dropped into
502 the tube, it stuck to the magnet and blocked access to the food regardless of how many light objects
503 were dropped. Thus, grackles had to inhibit dropping any heavy objects to solve this problem,
504 which made the task difficult. Despite the challenging apparatus, Refresco and Tequila likely would
505 have further changed their preference to light objects if given more trials since their performance
506 improved with the number of trials given, indicating that they were learning about the functional
507 properties of the task. New Caledonian crows previously showed behavioral flexibility on the
508 Narrow vs. Wide experiments when four crows preferred to drop objects into the functional narrow
509 tube rather than the non-functional wide tube, and then three changed their preference to the wide
510 tube and one changed to no preference when the wide tube became the functional option (Table 3,
511 Logan et al. 2014).

512 Contrary to the only previous study comparing species with different brain sizes, which
513 found a link between innovation frequency and brain size (Tebbich et al. 2010), I found no evidence
514 to validate the link between innovation and behavioral flexibility or the link between behavioral
515 flexibility and relative brain size when comparing non-innovative, average-brained great-tailed
516 grackles with innovative, large-brained New Caledonian crows. Both species exhibited behavioral
517 flexibility despite their differences in innovation frequency and relative brain size (Logan et al.
518 2014). My results are consistent with findings from the two studies on large-brained species that
519 directly investigated the relationship between innovation frequency and behavioral flexibility, and
520 behavioral flexibility and relative brain size (Auersperg et al. 2011; Manrique et al. 2013).

521 It could be argued that species that have many innovations are always behaviorally flexible,
522 but behavioral flexibility is not limited to species with many innovations. This would indicate

523 multiple causes of behavioral flexibility and could account for the small amount of variance
524 explained by the relationship between innovation frequency and relative brain size across species
525 (Lefebvre et al. 1997; Nicolakakis & Lefebvre 2000; Timmermans et al. 2000; Reader & Laland
526 2002; Overington et al. 2009). However, results from Manrique and colleagues (2013) suggest that
527 this is not the case: orang-utans, a species with many reported innovations, were less behaviorally
528 flexible than three other primate species with similar numbers of innovations. This suggests that
529 there is no causal link between behavioral flexibility and innovation frequency, though further
530 validations of this relationship would be useful.

531 Based on results from attempts to understand the relationship of behavioral flexibility to
532 innovation frequency and relative brain size, which now includes a sample of 13 species of birds
533 and mammals, it appears that behavioral flexibility should be quantified directly in each species
534 rather than using innovation frequency as an indirect proxy since it is unclear what innovation
535 frequency actually measures. Future research using proxies for behavioral flexibility at a broad
536 taxonomic scale should choose a proxy other than innovation frequency and validate it across a
537 number of species before relying on it. As the field stands now, it is unclear what the cross-species
538 correlations between innovation frequency and other factors imply.

539

540 **ACKNOWLEDGEMENTS**

541 I am grateful to Luisa Bergeron, Christin Palmstrom, Linnea Palmstrom, and Michelle Gertsvolf for
542 trapping and aviary assistance; Alexis Breen for helping to train Jugo; Brigit Harvey for stone
543 dropping training consultations; Steve Rothstein for scouting grackles and for use of the aviaries;
544 Joe Jablonski and David Bothman for making the apparatuses; Jill Zachary and Kathy Frye at Santa
545 Barbara City Parks and Recreation for use of the Andree Clark Bird Refuge and East Beach Park;
546 Estelle Sandhaus and Chris Briggs at the Santa Barbara Zoo for access to wild grackles; Karrie
547 Black for managing purchasing and grants; Alex Thornton for logistical advice; Dieter Lukas for
548 conceptual input and analysis feedback; Krista Fahy at the Santa Barbara Museum of Natural

549 History for helping with permit applications; Kristine Johnson, Sarah Overington, Julie Morand-
550 Ferron, and Neeltje Boogert for grackle advice and trap plans; Manny Garcia for veterinary and
551 permit support; Mary Hunsicker and Bertrand Lemasson for assistance with making the trap;
552 Margaret Tarampi, Eric Egenolf, Rebecca Schaefer, and Sam Franklin for brainstorming object
553 designs; Andrew Iwaniuk for access to the brain size database; Will Hoppitt for GLMM effect size
554 interpretation assistance; Irina Mikhalevich and Ljerka Ostojić for manuscript feedback; and my
555 funders: the SAGE Center for the Study of the Mind at the University of California Santa Barbara
556 and the National Geographic Society/Waite Grants Program.

557

558 REFERENCES

559 Akaike H (1981) Likelihood of a model and information criteria. *J Econometrics* 16: 3-14.

560 doi: 10.1016/0304-4076(81)90071-3

561 Alma S, Bee De Speroni N (1992) Indices cerebrales y composicion cuantitativa encefalica

562 en *Athene cunicularia* y *Tyto alba*. *Facena* 9:19-37.

563 Auersperg AMI, von Bayern AMP, Gajdon GK, Huber L, Kacelnik A (2011) Flexibility in

564 problem solving and tool use of kea and New Caledonian crows in a multi access box

565 paradigm. *PLOS ONE* 6:e20231. doi: 10.1371/journal.pone.0020231

566 Bates D, Maechler M, Bolker B (2011) lme4: Linear mixed-effects models using S4 classes. R

567 package version 0.999375-42. <http://CRAN.R-project.org/package=lme4>. Accessed 12

568 January 2015

569 Bee de Speroni N, Carezzano F (1995) Volumetric analysis of the visual, trigeminal and acoustic

570 nuclei in four avian species (Rheidae, Spheniscidae, Tinamidae). *Mar Ornithol* 23: 11-15.

571 Bird CD, Emery NJ (2009) Rooks use stones to raise the water level to reach a floating worm. *Curr*

572 *Biol* 19:1410-1414. doi: 10.1016/j.cub.2009.07.033

573 Boire D, Baron G (1994) Allometric comparison of brain and main brain subdivisions in birds.

574 *Journal of Brain Research* 35:49-66

- 575 Bond AB, Kamil AC, Balda RP (2007) Serial reversal learning and the evolution of behavioral
576 flexibility in three species of North American corvids (*Gymnorhinus cyanocephalus*,
577 *Nucifraga columbiana*, *Aphelocoma californica*). Journal of Comparative Psychology
578 121:372
- 579 Buckner C (2013) A property cluster theory of cognition. Philosophical Psychology 28:307-336.
580 doi: 10.1080/09515089.2013.843274
- 581 Burnham KP, Anderson DR (2002) Model selection and multimodel inference: a practical
582 information- theoretic approach, 2nd edn. Springer, New York.
- 583 Carezzano FJ, Bee De Speroni N (1995) Composicion volumetrica encefalica e indices cerebrales
584 en tres aves de ambiente acuatico (Ardeidae, Podicipedidae, Rallidae). Facena 11:75-83.
- 585 Cheke LG, Bird CD, Clayton NS (2011) Tool-use and instrumental learning in the Eurasian jay
586 (*Garrulus glandarius*). Anim Cogn 14:441-455. doi: 10.1007/s10071-011-0379-4
- 587 Cheke LG, Loissel E, Clayton NS (2012) How do children solve Aesop's Fable? PLOS ONE
588 7:e40574. doi: 10.1371/journal.pone.0040574
- 589 Crile G, Quiring DP (1940) A record of the body weight and certain organ and gland weights of
590 3690 animals. Ohio J Sci 40:219-259.
- 591 Day LB, Westcott DA, Olster DH (2005) Evolution of bower complexity and cerebellum size in
592 bowerbirds. Brain Behav Evol 66: 62-72.
- 593 Ducatez S, Clavel J, Lefebvre L (2015) Ecological generalism and behavioural innovation in birds:
594 technical intelligence or the simple incorporation of new foods? J Anim Ecol 84:79-89. doi:
595 10.1111/1365-2656.12255
- 596 Ebinger P (1995) Domestication and plasticity of brain organization in mallards (*Anas*
597 *platyrhynchos*). Brain Behav Evol 45: 286-300. doi: 10.1159/000113557
- 598 Fernandez P, Carezzano F, Bee De Speroni N (1997) Analisis cuantitativo encefalico e indices
599 cerebrales en *Aratinga acuticaudata* y *Myiopsitta monachus* de Argentina (Aves:
600 Psittacidae). Rev Chil Hist Nat 70:269-275.

- 601 Frahm HD, Rehkamper G (1998) Allometric comparison of the brain and brain structures in the
602 white crested Polish chicken with uncrested domestic chicken breeds. *Brain Behav Evol* 52:
603 292-307. doi: 10.1159/000006574
- 604 Ghahremani DG, Monterosso J, Jentsch JD, Bilder RM, Poldrack RA (2010) Neural components
605 underlying behavioral flexibility in human reversal learning. *Cerebral Cortex* 20:1843-1852
- 606 Griffin AS, Guez D (2015) Innovative problem solving in nonhuman animals: the effects of group
607 size revisited. *Behavioral Ecology* 26:722-734
- 608 Hadfield J (2010) MCMCglmm: Markov chain Monte Carlo methods for generalised linear mixed
609 models.
610 [http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=E395FF7367BF580D916E465F55](http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=E395FF7367BF580D916E465F55302C60?doi=10.1.1.160.5098&rep=rep1&type=pdf)
611 [302C60?doi=10.1.1.160.5098&rep=rep1&type=pdf](http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=E395FF7367BF580D916E465F55302C60?doi=10.1.1.160.5098&rep=rep1&type=pdf) Accessed 7 May 2015
- 612 Hadfield J (2014) MCMCglmm course notes. [http://cran.r-](http://cran.r-project.org/web/packages/MCMCglmm/vignettes/CourseNotes.pdf)
613 [project.org/web/packages/MCMCglmm/vignettes/CourseNotes.pdf](http://cran.r-project.org/web/packages/MCMCglmm/vignettes/CourseNotes.pdf). Accessed 7 May 2015
- 614 Hadfield JD, Nakagawa S (2010) General quantitative genetic methods for comparative biology:
615 Phylogenies, taxonomies, meta-analysis and multi-trait models for continuous and
616 categorical characters. *J Evol Biol* 23:494–508.
- 617 Healy SD, Rowe C (2007) A critique of comparative studies of brain size. *Proc R Soc B* 274:453-
618 464. doi: 10.1098/rspb.2006.3748
- 619 Hunt GR (1996) Manufacture and use of hook-tools by New Caledonian crows. *Nature* 379:249-
620 251. doi: 10.1038/379249a0
- 621 Hunt GR (2000) Tool use by the New Caledonian crow *Corvus moneduloides* to obtain
622 Cerambycidae from dead wood. *Emu* 100:109-114
- 623 Hunt GR (2008) Introduced *Lantana camara* used as tools by New Caledonian crows (*Corvus*
624 *moneduloides*). *N Z J Zool* 35:115-118.
- 625 Hunt GR, Gray RD (2002) Species-wide manufacture of stick-type tools by New Caledonian crows.
626 *Emu* 102:349-353

- 627 Hunt GR, Gray RD (2004) Direct observations of pandanus-tool manufacture and use by a New
628 Caledonian crow (*Corvus moneduloides*). *Anim Cogn* 7:114-120. doi: 10.1007/s10071-003-
629 0200-0
- 630 Iwaniuk AN (2004) Brood parasitism and brain size in cuckoos: A cautionary tale on the use of
631 modern comparative methods. *Int J Comp Psychol* 17: 17-33.
- 632 Iwaniuk AN, Dean KM, Nelson JE (2005) Interspecific allometry of the brain and brain regions in
633 parrots (Psittaciformes): comparisons with other birds and primates. *Brain Behav Evol*
634 65:40-59. doi: 10.1159/000081110
- 635 Iwaniuk AN, Hurd PL (2005) The evolution of cerebrotypes in birds. *Brain Behav Evol* 65:215-
636 230. doi: 10.1159/000084313
- 637 Iwaniuk AN, Nelson JE (2001) A comparative analysis of relative brain size in waterfowl
638 (Anseriformes). *Brain Behav Evol* 57:87-97. doi: 10.1159/000047228
- 639 Iwaniuk AN, Nelson JE (2003) Developmental differences are correlated with relative brain size in
640 birds: a comparative analysis. *Can J Zool* 81: 1913-1928.
- 641 Iwaniuk AN, Nelson JE, James HF, Olson SL (2004) A comparative test of the correlated evolution
642 of flightlessness and relative brain size in birds. *J Zool Lond* 263: 317-327. doi:
643 10.1017/S0952836904005308
- 644 Iwaniuk AN, Wylie DRW (2006) The evolution of stereopsis and the Wulst in caprimuliform
645 birds: a comparative analysis. *J Comp Physiol A* 192: 1313-1326. doi: 10.1007/s00359-006-
646 0161-2
- 647 Jelbert SA, Taylor AH, Cheke LG, Clayton NS, Gray, RD (2014) Using the Aesop's fable paradigm
648 to investigate causal understanding of water displacement by New Caledonian crows. *PLOS*
649 *ONE* 9:e92895. doi: 10.1371/journal.pone.0092895
- 650 Jerison HJ (1985) Animal intelligence as encephalization. *Phil Trans R Soc Lond B* 308:21-35
- 651 Layard EL, Layard ELC (1882) Notes on the avifauna of New Caledonia. *Ibis* 6:520-522.

- 652 Lefebvre L, Whittle P, Lascaris E, Finkelstein A (1997) Feeding innovations and forebrain size in
653 birds. *Anim Behav* 53:549-560. doi: 10.1006/anbe.1996.0330
- 654 Lefebvre L, Reader SM, Sol D (2004) Brains, innovation and evolution in birds and primates. *Brain*
655 *Behav Evol* 63:233-246. doi: 10.1159/000076784
- 656 Lefebvre L, Reader SM, Sol D (2013) Innovating innovation rate and its relationship with brains,
657 ecology and general intelligence. *Brain Behav Evol* 81:143-145. doi: 10.1159/000348485
- 658 Le Goupils, M., 1928. Dans la Brousse Calédonienne: Souvenirs d'un Ancien Planteur, 1898-1904.
659 Perrin, Paris.
- 660 Logan CJ, Jelbert SA, Breen AJ, Gray RD, Taylor AH (2014) Modifications to the Aesop's Fable
661 paradigm change performances in New Caledonian crows. *PLOS ONE* 9:e103049.
662 doi:10.1371/journal.pone.0103049
- 663 Logan C (2015) Great-tailed grackle water tube experiments, Santa Barbara, CA USA 2014-2015.
664 KNB Data Repository. doi:
- 665 Manrique HM, Völter CJ, Call J (2013) Repeated innovation in great apes. *Anim Behav* 85:195-
666 202. doi: 10.1016/j.anbehav.2012.10.026
- 667 Mlikovsky J (1989a) Brain Size in Birds: 1. Tinamiformes Through Ciconiformes, *Vestrik*
668 *Ceskoslovenska Spolecnosti Zoologicka* 53:33-47.
- 669 Mlikovsky J (1989b) Brain Size in Birds: 2. Falconiformes through Gaviiformes. *Vestrik*
670 *Ceskoslovenska Spolecnosti Zoologicka* 53:200-213.
- 671 Mlikovsky J (1989c) Brain Size in Birds: 3. Columbiformes Through Piciformes. *Vestrik*
672 *Ceskoslovenska Spolecnosti Zoologicka* 53 252-273.
- 673 Mlikovsky J (1990) Brain Size in Birds: 4. Passeriformes. *Acta Soc Zool Bohemoslov* 54:27-101.
- 674 Møller AP, Erritzoe J, Garamszegi LZ (2004) Covariation between brain size and immunity in
675 birds: implications for brain size evolution. *J Evol Biol* 18:223-237. doi: 10.1111/j.1420-
676 9101.2004.00805.x

- 677 Nicolakakis N, Lefebvre L (2000) Forebrain size and innovation rate in European birds: feeding,
678 nesting and confounding variables. *Behav* 137:1415-1429. doi: 10.1163/156853900502646
- 679 Overington SE, Morand-Ferron J, Boogert NJ, Lefebvre L (2009) Technical innovations drive the
680 relationship between innovativeness and residual brain size in birds. *Anim Behav* 78:1001-
681 1010. doi: 10.1016/j.anbehav.2009.06.033
- 682 Overington SE, Cauchard L, Cote KA, Lefebvre L (2011) Innovative foraging behavior in birds:
683 What characterizes an innovator? *Behav Process* 87:274-285. doi:
684 10.1016/j.beproc.2011.06.002
- 685 Pistone E, Carezzano F, Bee De Speroni N (2002) Tamano relativo encefalico e indices cerebrales
686 en *Vanellus c. chilensis* (Aves: Charadriidae). *Rev Chil Hist Nat* 75: 595-602. doi:
687 10.4067/S0716-078X2002000300010
- 688 Pyle P (2001) Identification guide to North American birds part 1. Sheridan Books, Inc., Ann
689 Arbor, Michigan.
- 690 R Core Team (2014) R: A language and environment for statistical computing. R Foundation for
691 Statistical Computing, Vienna, Austria. <http://www.R-project.org>. Accessed 3 Jan 2015.
- 692 Reader SM, Laland KN (2002) Social intelligence, innovation, and enhanced brain size in primates.
693 *Proceedings of the National Academy of Sciences* 99:4436-4441. doi:
694 10.1073/pnas.062041299
- 695 Rehkämper G, Frahm HD, Cnotka J (2008) Mosaic evolution and adaptive brain component
696 alteration under domestication seen on the background of evolutionary theory. *Brain Behav*
697 *Evol* 71: 115-126. doi: 10.1159/000111458
- 698 Rehkämper G, Frahm HD, Zilles K (1991) Quantitative development of brain and brain structures
699 in birds (Galliformes and Passeriformes) compared to that in mammals (Insectivores and
700 Primates). *Brain Behav Evol* 37:125-143. doi: 10.1159/000316088
- 701 Rehkämper G, Kart E, Frahm HD, Werner CW (2003) Discontinuous variability of brain
702 composition among domestic chicken breeds. *Brain Behav Evol* 61: 59-69.

- 703 Rutz C, Bluff LA, Weir AAS, Kacelnik A (2007) Video cameras on wild birds. *Science* 318:765
- 704 Taylor AH, Elliffe DM, Hunt GR, Emery NJ, Clayton NS, Gray RD (2011) New Caledonian crows
705 learn the functional properties of novel tool types. *PLOS ONE* 6:e26887.
706 doi:10.1371/journal.pone.0026887
- 707 Tebbich S, Sterelny K, Teschke I (2010) The tale of the finch: adaptive radiation and behavioural
708 flexibility. *Philos Trans R Soc B* 365:1099-1109. doi: 10.1098/rstb.2009.0291
- 709 Timmermans S, Lefebvre L, Boire D, Basu P (2000) Relative size of the hyperstriatum ventrale is
710 the best predictor of feeding innovation rate in birds. *Brain Behav Evol* 56:196-203. doi:
711 10.1159/000047204
- 712 Troscianko J, Bluff LA, Rutz C (2008) Grass-stem tool use in New Caledonian Crows *Corvus*
713 *moneduloides*. *Ardea* 96:283-285