

Wild psychometrics: evidence for “general” cognitive performance in wild New Zealand robins *Petroica longipes*

Rachael C. Shaw^{a*}, Neeltje J. Boogert^b, Nicola S. Clayton^c, Kevin C. Burns^a

^a School of Biological Sciences, Victoria University of Wellington, New Zealand

^b School of Psychology and Neuroscience, University of St Andrews, United Kingdom

^c Department of Experimental Psychology, University of Cambridge, United Kingdom

Word count: 7 775 (excluding title page, references, figures and tables)

* Correspondence:

Postal: R Shaw, School of Biological Sciences, Victoria University of Wellington, PO Box 600,
Wellington 6140, New Zealand

Email: rachael.c.shaw@gmail.com

Phone: +64 (0)21 236 8403

1 **ABSTRACT**

2 In human psychometric testing, individuals' scores in tests of diverse cognitive processes are
3 positively correlated, with a "general intelligence" factor (*g*) typically accounting for at least 40%
4 of total variance. Individual differences in cognitive ability have been extensively studied in
5 humans, yet they have received far less attention in non-human animals. In particular, the
6 development of a test battery suitable for quantifying individual cognitive performance in birds
7 remains in its infancy. Additionally, implementing this approach in the wild, where the ecological
8 significance of cognition can also be explored, presents considerable logistical challenges for most
9 species. We developed a cognitive test battery for wild New Zealand North Island robins (*Petroica*
10 *longipes*). Our battery was comprised of six tasks based on established measures of avian cognitive
11 performance: a motor task, colour and shape discrimination, reversal learning, spatial memory and
12 inhibitory control. Robins ($N = 20$) varied greatly in their ability to solve these tasks and we found
13 weakly positive, non-significant correlations between most tasks. A principal components analysis
14 (PCA) of task performances yielded two factors with eigenvalues >1 . The first component extracted
15 explained over 34% of the variance in cognitive performance and all six tasks loaded positively on
16 this first component (mean loading \pm SD = 0.559 ± 0.196). We show that these results are robust
17 using randomisation tests. Our results thus suggest that a general cognitive factor, analogous to
18 human *g*, underpins cognitive performance in wild North Island robins tested in their natural
19 habitat.

20

21 **Keywords:** cognitive test battery, *g*, general intelligence factor, individual variation, *Petroica*
22 *longipes*

23 **INTRODUCTION**

24 Research interest in non-human animal cognition – the mechanisms by which individuals acquire,
25 process, store and act on information in their environment (Shettleworth, 2010) – has grown
26 substantially in recent years. Inter-individual variation in cognitive performance was previously
27 perceived as uninteresting ‘noise’ around the mean, however, researchers are increasingly
28 recognizing that such individual differences can have important consequences for survival and
29 reproduction (e.g. Cauchard, Boogert, Lefebvre, Dubois, & Doligez, 2013; Cole, Morand-Ferron,
30 Hinks, & Quinn, 2012; Isden, Panayi, Dingle, & Madden, 2013; Keagy, Savard, & Borgia, 2009,
31 2011). If we are to assess the fitness consequences of cognitive ability, we first need robust and
32 ecologically relevant measures of individual cognitive variation (Rowe & Healy, 2014). Recent
33 research has largely focussed on “novel problem-solving performance” (reviewed in Thornton,
34 Isden, & Madden, 2014). In this approach, animals are presented with a single task, such as pulling
35 a lever to release a food reward (Cole et al., 2012), removing an obstruction from their nest box
36 entrance (Cauchard et al., 2013), or removing undesirable objects from a display bower (Keagy et
37 al., 2011, 2009). Those individuals that manage to solve the task, or solve the task faster than
38 others, are deemed to have “better” cognitive ability (Rowe & Healy, 2014; Thornton et al., 2014).
39 However, it remains unclear exactly which cognitive abilities such tasks are actually measuring.
40 Furthermore, these tasks are often presented only once to each test subject. A one-off task solve
41 may be due to chance or a combination of non-cognitive factors, such as persistence, motivation or
42 dexterity (Thornton et al., 2014). Instead, the use of test batteries that target defined cognitive
43 processes has recently been advocated (Isden et al., 2013; Thornton et al., 2014; Thornton, 2014).

44
45 In human psychometric studies, individuals’ scores in test batteries assessing diverse cognitive
46 processes, such as processing speed, working memory and verbal comprehension, are positively
47 correlated, with a single factor – termed *g* (for “general intelligence”) – typically accounting for at
48 least 40% of the total variance (Deary, Penke, & Johnson, 2010; Deary, Spinath, & Bates, 2006;
49 Plomin & Spinath, 2002; Plomin, 2001). In non-human animals, positive correlations between

50 performance scores on tasks assessing different types of learning have been documented in
51 honeybees (*Apis mellifera*; Chandra, Hosler, & Smith, 2000) and feral pigeons (*Columba livia*;
52 Bouchard, Goodyer, & Lefebvre, 2007), while cognitive test batteries have been developed to test
53 for *g* in several primate species (Amici, Barney, Johnson, Call, & Aureli, 2012; Banerjee et al.,
54 2009; Herrmann, Call, Hernández-Lloreda, Hare, & Tomasello, 2007; Schmitt, Pankau, & Fischer,
55 2012) and in mice (*Mus musculus*; Galsworthy et al., 2005; Locurto, Fortin, & Sullivan, 2003;
56 Matzel et al., 2003). However, the specific tasks used in a test battery may have implications for
57 interpreting correlations between performance scores. For example, the mouse test batteries
58 typically contain a predominance of spatial tasks. Positive correlations may thus arise because most
59 tests are tapping into the same cognitive process (e.g. spatial learning/memory; Amici et al., 2012).
60 Reliably estimating *g* therefore requires choosing tasks which span different cognitive domains
61 (Amici et al., 2012; Thornton et al., 2014).

62

63 To date there have been few attempts to develop avian cognitive test batteries. Boogert and
64 colleagues (Boogert, Anderson, Peters, Searcy, & Nowicki, 2011) tested wild-caught song sparrows
65 (*Melospiza melodia*) in the laboratory using four tasks (a motor task, colour discrimination, colour
66 reversal and a detour-reaching task). Isden et al. (2013) gave male spotted bowerbirds
67 (*Ptilonorhynchus maculatus*) a test battery consisting of six tasks (a problem-solving task, a motor
68 task, colour discrimination, colour reversal, shape discrimination and spatial memory) and a general
69 factor was found to account for over 44% of the variance in task performance (Isden et al., 2013).
70 These pioneering avian test batteries did not contain as many tasks as those used for primates (e.g.
71 Amici et al., 2012; Banerjee et al., 2009; Herrmann et al., 2007; Schmitt et al., 2012a) and mice
72 (Galsworthy et al., 2005; Locurto et al., 2003; Matzel et al., 2003). However, the tasks that they
73 incorporated were based on established, psychologically-grounded experimental methods for
74 assessing defined cognitive traits (Boogert, Anderson, et al., 2011; Isden et al., 2013), a
75 characteristic that is critical in the design of cognitive test batteries (Thornton et al., 2014).

76

77 Most animal cognition studies to date have tested small groups of captive individuals. However, this
78 approach overlooks the ecological and evolutionary significance of cognition (Thornton & Lukas,
79 2012) and may generate data that are confounded by inter-individual differences in response to
80 captivity. Unfortunately, getting wild birds to participate in a series of cognitive tasks in the field is
81 prohibitively challenging in most species (Boogert, Fawcett, & Lefebvre, 2011; Boogert, Monceau,
82 & Lefebvre, 2010). Implementing a battery of tasks in the field requires that an animal can be
83 located over multiple testing sessions and is willing to approach and interact with novel objects,
84 without interference from conspecifics. As yet, few species have been identified that fulfil all these
85 requirements (Isden et al., 2013; Keagy et al., 2011).

86
87 The North Island robin (*Petroica longipes*), a small insectivorous passerine that is endemic to New
88 Zealand, is an ideal species for implementing a cognitive test battery in the field. North Island
89 robins lack neophobia and have few anti-predatory behaviours (Maloney & Mclean, 1995). They
90 readily participate in cognition experiments requiring them to interact with humans and novel
91 objects, including tests of quantity discrimination (Garland, Low, & Burns, 2012; Hunt, Low, &
92 Burns, 2008) and human gaze avoidance (Garland, Low, Armstrong, & Burns, 2014). Moreover,
93 robins are territorial year-round, meaning that individuals can be reliably located for participation in
94 multiple testing sessions. North Island robins are also a food-hoarding species that will dismember
95 large insect prey and store them in branch-trunk axils in the forest canopy (Steer & Burns, 2008;
96 Van Horik & Burns, 2007). The spatial cognition underpinning food-hoarding behaviour has been
97 investigated in several species (reviewed in Smulders, Gould, & Leaver, 2010) and many other
98 cognitive domains have been tested in food-hoarding species. For example, Eurasian jays, *Garrulus*
99 *glandarius*, have been given tasks involving instrumental learning (Cheke, Bird, & Clayton, 2011),
100 inference (Shaw, Plotnik, & Clayton, 2013), social cognition (Ostojić, Shaw, Cheke, & Clayton,
101 2013; Shaw & Clayton, 2013), future planning (Cheke & Clayton, 2012) and inhibitory control
102 (MacLean et al., 2014). However, a cognitive test battery has not yet been conducted to explicitly
103 test for *g* in any food-hoarding species. The North Island robin therefore provides the opportunity to

104 test for the presence of a factor analogous to human *g* in a species that is highly likely to have
105 undergone selection to specialize in the spatial cognitive domain.

106

107 The aim of this study was to quantify individual cognitive variation of wild North Island robins and
108 determine whether this variation could be described with a single general cognition factor (i.e. *g*). **In**
109 **contrast to previous non-human studies, we also tested whether our results were robust using**
110 **randomisation tests.** Our test battery was **comprised of six psychologically-grounded tasks:** a motor
111 task, colour discrimination and reversal learning, shape discrimination, spatial memory and detour-
112 reaching.

113

114 **METHODS**

115 *Study Site and Subjects*

116 We conducted the experiments between 3 April and 25 September 2014 at Zealandia Wildlife
117 Sanctuary, a 225 ha wildlife sanctuary in Central Wellington (41°18'S, 174°44'E) that is
118 surrounded by a predator exclusion fence. Since 2000 all mammalian predators have been removed
119 from the reserve (with the exception of mice) and many rare forest-dwelling endemic birds have
120 been reintroduced to the site. In 2001 and 2002, a total of 76 North Island robins were translocated
121 from Kapiti Island to Zealandia and the population has since been breeding successfully, with robin
122 density estimated to be between 2.3 and 3.4 individuals per ha in 2008 (i.e. between 500 and 765
123 birds in total; McGavin, 2009).

124

125 Male robins are typically dominant to their mates and will displace females from food sources
126 (Burns and Steer, 2006). The majority of our 20 adult subjects were therefore male (males = 14,
127 females = 4, sex unknown = 2). Our research was conducted within a 25 ha area of the sanctuary.
128 We individually banded robins with 3 plastic coloured bands and a metal band (2 bands per leg,
129 **bands supplied by the Department of Conservation's National Banding Office).** Robins were caught
130 **using a drop trap and released immediately after banding, weighing, tarsus and wing chord**

131 measurements were complete. To ensure there were no adverse effects of the banding procedures
132 robins were fed two mealworms immediately after release and were frequently monitored. There
133 was no evidence of banding related injuries during the experiment. The exact age of the test
134 subjects was unknown.

135

136 ***Ethical Note***

137 The research was approved by the Victoria University of Wellington Animal Ethics Committee and
138 conducted under permit from the Department of Conservation (Authorisation number: 38497-FAU).

139

140 ***Cognitive Test Battery***

141 We ran experiments between 0830 and 1430. All robins participated voluntarily in the cognitive
142 experiments. The cognitive test battery consisted of six tasks with a consistent task order to
143 standardize carry-over effects: 1) motor task, 2) colour discrimination, 3) colour reversal, 4) spatial
144 memory, 5) detour reaching and 6) symbol discrimination. Three subjects disappeared before
145 completing the final symbol discrimination task (one was displaced from her winter territory mid-
146 way through the task and two were either predated or displaced before beginning the task).

147

148 A wooden board (25 × 35 cm) served as a testing platform for all tasks (Fig. 1). We placed the
149 platform on the ground under a section of closed canopy within a bird's territory, at least 8 m from a
150 territory boundary. We identified boundaries during territorial displays between the territory holder
151 and any neighbouring robins. The location and orientation of the platform, as well as the placement
152 of any apparatus on it was consistent across all trials. All trials were observed and scored live by the
153 experimenter (RCS) who stood at least 1.5 m from the platform. Trials were also filmed with a
154 Sony HDR-AS30V camera to check scores later. To avoid territorial disputes and the possibility of
155 social learning, a trial began when the robin was in view of the experimenter and no conspecifics
156 could be seen or heard in the area.

157

158 Tasks 1-4 and 6 used foraging grids consisting of a wooden block ($13.5 \times 9 \times 2.5$ cm) containing 6
159 wells (1.3 cm diameter, 0.8 cm deep; Fig. 1a-c). The wells could be covered with circular
160 removable PVC lids (2.2 cm diameter, 0.05 cm high). A vinyl bumper (1.3 cm diameter, 0.3 cm
161 high) was glued to the bottom of each lid and fitted exactly into the well (as in Boogert et al.,
162 2011a).

163

164 ***Motor Task***

165 To assess motor skill learning we adapted protocols that have been used with captive, domesticated
166 zebra finches (*Taeniopygia gutatta*; Boogert et al., 2008) and wild-caught song sparrows (Boogert
167 et al., 2011a). We trained robins to flip white PVC lids on the foraging grid using a systematic
168 shaping procedure with four training stages, following Boogert et al. (2008) (Table 1). In each trial,
169 we placed one foraging grid on the testing platform and baited four of its six wells with a freshly
170 killed mealworm (*Tenebrio molitor* larvae; Fig. 1a). We allowed robins up to 3 min to retrieve
171 mealworms before removing the grid, as this was the maximum amount of time birds needed to
172 retrieve at least 3 out of 4 mealworms during level 1. Between trials there was a delay of 1-3 min
173 during which we recorded the previous trial outcome and reset the apparatus. We kept the inter-trial
174 interval to a minimum to ensure that a robin did not leave the immediate area during a test session.
175 We randomised the location of the baited wells between trials. Robins were presented with one test
176 session, consisting of five trials, per day.

177

178 We considered a robin to have solved the task once it retrieved at least three mealworms in 6 out of
179 7 consecutive trials (across sessions) when the lids were fully inserted into the wells (level 4 pass
180 criterion, Table 1). If a robin retrieved no mealworms in three consecutive trials, it regressed to the
181 previous level. The motor task performance measure that we used in the analyses was the summed
182 number of trials the robin took to pass both levels 3 and 4 of the task, when the reward was not
183 visible (following Boogert et al., 2011a).

184

185 *Colour Preference and Colour Discrimination*

186 After a robin had completed level 4 of the motor task, we gave it a colour preference test on the
187 following test day (which did not always fall on the following calendar day). We presented the
188 robin with a single foraging grid containing two wells side-by-side that were baited with a freshly
189 killed mealworm and were covered with differently coloured lids (Fig. 1b). We allowed the robin to
190 flip one lid only and this was taken to be their preferred colour.

191
192 Out of the 20 test subjects, 12 robins were presented with one red lid and one blue lid for the
193 preference test. However, eight other robins had previously participated in a pilot test of a different
194 colour association task using red and blue lids. Although none of the robins fulfilled the learning
195 criteria in that pilot test, we presented these eight robins with green and yellow lids to minimise any
196 carry-over effects from the pilot test.

197
198 Immediately after the colour preference test we gave a robin their first session of the colour
199 association task. For this, we presented the robin with a single foraging grid with one red lid and
200 one blue lid (or one yellow and one green for the previously tested group) covering two wells that
201 were side-by-side. Whichever colour was not flipped first by the robin in the preference test now
202 covered a well baited with one freshly killed mealworm. The preferred coloured lid covered a well
203 that was empty. The first trial was a probe trial where we allowed robins to flip both lids, so that
204 they discovered that only one of the wells contained a mealworm. In all remaining test trials we
205 allowed a robin to flip only one lid before the experimenter removed the apparatus, so that
206 removing the un-rewarded lid came at the cost of a missed mealworm reward. Robins had 2 min to
207 complete a trial, with a 1-3 min delay between trials.

208
209 We pseudo-randomised the side of the apparatus containing the baited well between test trials, such
210 that one side of the grid was never baited in more than three consecutive trials. Additionally, we
211 never used the same pair of wells between consecutive trials, to minimise the possibility of robins

212 experiencing interference from spatial cues remembered from the preceding trial. If a robin failed to
213 retrieve any worms in five consecutive trials, we gave it a single worm to prevent it from becoming
214 so hungry that it began foraging for naturally occurring foods.

215
216 We gave robins one session per day, consisting of 15 test trials. A robin solved the task if it flipped
217 the rewarded colour in 10 out of 12 consecutive trials, either within a single session or across two
218 test sessions (as this exceeds the chance expectation of 6 out of 12 trials correct: two-tailed binomial
219 test $P = 0.039$). Unless the weather was too inclement for testing to be carried out safely, we
220 conducted sessions on consecutive days until the robin solved the task.

221

222 ***Colour Reversal***

223 Reversal learning tasks have been used to measure behavioural flexibility in many avian species
224 (e.g. corvids: Bond et al., 2007; Zenaida doves, *Zenaida aurita*: Boogert et al., 2010; Darwin's
225 finches: Tebbich et al., 2010). The day after a robin solved the colour discrimination task, we
226 presented the same task but reversed the colour of lid that hid the mealworm (e.g. blue now hid the
227 mealworm if red had covered it in the colour discrimination task). The procedure was otherwise
228 identical to the colour association task, with the pass criterion being to flip the rewarded colour in
229 10 out of 12 consecutive trials.

230

231 ***Spatial Memory***

232 Avian spatial memory studies typically investigate how quickly a bird learns to use spatial cues to
233 accurately locate a food in a foraging grid (e.g. western scrub-jays, *Aphelocoma californica*:
234 Pravosudov, Lavenex, & Omanska, 2005; zebra finches: Sanford & Clayton, 2008; song sparrows:
235 Sewall, Soha, Peters, & Nowicki, 2013). Our spatial memory task was a simplified and shortened
236 version of a protocol previously used to test spatial memory in wild-caught song sparrows (Sewall
237 et al., 2013). We gave robins two foraging grids placed side by side to create a 3 × 4 grid (Fig. 1c).
238 Eight of the 12 wells were covered with grey lids (avoiding the corner wells). Only one of the eight

239 wells contained a freshly killed mealworm. The exact location of the reward was randomised
240 between birds.

241

242 On the first day of testing we gave robins a baseline trial in which they could flip all lids to locate
243 the single food reward. We removed the grids and lids from the testing platform after all lids were
244 flipped. The birds had a training trial after a 5 min delay. We placed the two grids on the testing
245 platform in the same orientation with the same well containing the mealworm and allowed the robin
246 to flip all lids to search for the food reward. The following day, after a 24 h delay, we presented the
247 apparatus in the same location and orientation, with the same well baited. We repeated the
248 procedure again the following day, 48 h after the first trial. In every trial, we only removed the grid
249 once the robin had flipped all eight lids (following the procedures of Sewall et al., 2013).

250

251 We used the total number of lids that a robin flipped during the two test trials (24 h and 48 h delay)
252 before flipping the rewarded lid as a measure of its spatial memory performance (following Sewall
253 et al., 2013). To ascertain whether the robins could use odour cues, we conducted a final probe trial
254 5 minutes after the 48 h spatial memory test. In this probe, none of the wells were baited. We also
255 rotated the grid, so that its appearance did not change, but the well that had contained food in the
256 preceding trial was now on the opposite side of the grid. This ensured that there were no residual
257 odour cues at the correct spatial location.

258

259 ***Detour Reaching***

260 Inhibitory control (characterised as the ability to inhibit a pre-potent response) is a crucial
261 component of executive function and decision making in humans and non-human animals (Hauser,
262 1999; MacLean et al., 2014). It has been measured in several bird species with a detour-reaching
263 task in which an animal must learn to retrieve a reward from behind a transparent barrier without
264 first attempting to reach through this barrier (Boogert, Anderson, et al., 2011; MacLean et al.,
265 2014). For our detour-reaching task the apparatus was a transparent plastic cylinder (5 cm length, 4

266 cm diameter, 0.1 cm thickness, open at both ends and glued to a wooden base (5 × 4 cm and 0.6 cm
267 high; for a detailed picture of the apparatus see Fig. 1 in Boogert, Anderson, et al., 2011).

268

269 The task consisted of habituation, training and test phases. During the habituation and training
270 phases, the cylinder was opaque (wrapped in black electrical tape with both ends remaining open,
271 Fig. 1d). For habituation trials, we gave the robin the opaque tube with a single, freshly killed
272 mealworm placed in the centre. Habituation was complete when the robin consumed the mealworm
273 within 2 min in three consecutive trials. Training was identical to habituation, and was completed
274 when the robin removed the worm from the open end of the tube without pecking at the tube first in
275 four consecutive trials. The opaque tube-training phase ensured that the robin had learned to move
276 to the open end of the tube to retrieve the worm, so that the test phase only tested the robin's
277 inhibitory control performance. We oriented the tube so that the plastic side was parallel with the
278 platform side that the robin was most likely to approach first (we defined this as the side of the
279 platform that a robin approached first most frequently during all trials of the Spatial Memory task).

280

281 During the testing phase the tube was transparent (Fig. 1e). To pass, the robin had to successfully
282 detour to the open end of the tube to retrieve the worm, without pecking at the transparent wall of
283 the tube first, in six out of seven consecutive trials. The performance measure was how many trials
284 the robin took to reach this criterion. There was an interval of 1 min between trials and we gave
285 robins no more than 20 trials in a day.

286

287 *Symbol Discrimination*

288 The symbol discrimination task protocol was identical to the colour discrimination task. We initially
289 tested a robin's preference for two symbols (a cross and a square) that were matched in terms of the
290 area they covered on a white background and the amount of black line they contained. We then used
291 the symbol that was not flipped during the preference test to cover the well containing the food
292 reward in the test trials. The first 10 robins that we tested on this task received up to eight 15-trial

293 sessions (i.e. a total of 120 trials). However, to ensure that we completed all cognitive testing before
294 the onset of nesting, we gave the remaining seven birds three sessions (45 trials) only.

295

296 *Motivation and Neophobia*

297 To ensure that robins were motivated for food rewards throughout cognitive test sessions, they had
298 been trained before cognitive tests commenced to hop on an electronic scale for a single mealworm
299 reward, and were weighed before and after every **testing** session. Such motivation checks are
300 critical to interpret cognitive measures; motivational factors may not necessarily relate to the
301 cognitive abilities required to solve a task, but may greatly affect test performance (Rowe and
302 Healy, 2014; Sanford and Clayton, 2008). To characterise a robin's neophobia, when we gave a
303 robin a new apparatus (or a new lid colour) we recorded how long it took for the robin to touch the
304 item.

305

306 *Statistical Analyses*

307 Statistical analyses were conducted in R version 3.1.1 (The R Foundation for Statistical Computing,
308 Vienna, Austria, <http://www.r-project.org>). For all our performance measures, a lower score
309 indicates that a robin required fewer test trials to reach the task's pass criterion. **In the spatial
310 memory task we also calculated how many lids the robins were expected to flip if searching
311 randomly during a trial. For this we used an equation derived from the negative hypergeometric
312 distribution (assuming sampling without replacement; equation 8 in Tillé, Newman, & Healy,
313 1996). We used two-tailed, one sample t-tests to compare this random search expectation to the
314 observed number of lids flipped by robins in the 24 h and in the 48 h test trial.**

315

316 To explore how task performances in the cognitive test battery were related to each other we used
317 Spearman rank correlations to test for pairwise correlations. A Bonferroni correction for multiple
318 comparisons was applied to the alpha-level of significance. To investigate whether inter-individual
319 variation in performance across cognitive tasks could be explained by a single factor extracted from

320 the data (i.e. g) we performed a principle components analysis (PCA) with an unrotated factor
321 solution using the princomp function in R. The scores from the first unrotated principal component
322 are widely used as a measure of g in both humans and non-human animals (e.g. Isden et al., 2013;
323 Locurto et al., 2003; Plomin and Spinath, 2002; Plomin, 2001). All tasks loading positively onto the
324 first unrotated component that also explains 30-45% of the variance in test performance has
325 previously been interpreted as evidence for g in non-human studies of the structure of cognitive
326 performance (e.g. Galsworthy et al., 2005b; Isden et al., 2013; Matzel et al., 2003). To assess the
327 likelihood that all tasks would load positively on the first unrotated component extracted, we
328 compared our results to the results obtained in 10 000 simulations. For each simulation we
329 randomised the test battery scores between birds (using the randomizeMatrix function in R package
330 picante: Kembel et al., 2014), performed an unrotated PCA and obtained the mean and standard
331 deviation of the factor loadings for the first unrotated component extracted.

332

333 **RESULTS**

334 Robins ($N = 20$) required mean \pm SD = 17.20 ± 2.29 days (range 14-22) to complete the first five
335 tasks of the test battery.

336

337 ***Motor Task***

338 Four robins passed the final two levels of the task, in which the reward was completely covered by
339 the lid, in the minimum number of trials possible (10 trials). The slowest individual required 18
340 trials to pass the final two stages of the motor task (mean \pm SD trials to pass levels 3 and 4 = 13.00
341 ± 2.36 , $N = 20$).

342

343 ***Colour Preference and Colour Discrimination***

344 Of the twelve robins that were presented with one red and one blue lid, eight chose the blue lid first
345 in the preference test. Of the eight birds that were presented with green and yellow lids, four chose

346 the green lid first in the preference test. The number of trials that robins required to solve the colour
347 discrimination task ranged between 12 and 80 (mean \pm SD = 40.05 \pm 19.33, $N = 20$).

348

349 Individuals' colour discrimination performance was correlated with their lid colour preference
350 (Kruskal-Wallis test: $K_3 = 11.06$, $N = 20$, $P = 0.011$). The four robins that preferred the red lid (and
351 hence had blue lids rewarded in the test) required the fewest trials to reach the learning criterion
352 (Fig. 2).

353

354 *Colour Reversal*

355 Robins took longer to learn the reversal than they did to learn the original colour association (paired
356 t test: $t_{19} = -3.751$, $P = 0.001$), with the number of trials required to solve the task ranging between
357 33 and 89 (mean \pm SD = 58.60 \pm 15.54, $N = 20$). In contrast to the colour discrimination task, initial
358 colour lid preference did not affect how long it took a robin to learn the reversal (Kruskal-Wallis
359 test: $K_3 = 0.754$, $N = 20$, $P = 0.861$).

360

361 *Spatial Memory*

362 Robins made between 0 and 13 errors (i.e. incorrect lids flipped before finding the mealworm) in
363 total across the 24h and 48h memory tests (mean \pm SD = 7.32 \pm 3.25, $N = 19$). One robin cached
364 several lids during the first presentation of the task. As a result, there were insufficient lids
365 remaining to run his 5 min training trial and he was subsequently excluded from the experiment.

366

367 The random search expectation per trial was 4.5 lids (calculated following the methods of Tillé et
368 al., 1996). Overall the robins' search behaviour did not differ from the random search expectation in
369 either the 24h test (mean \pm SD = 4.58 \pm 2.14 one-sample t test: $t_{18} = 0.1606$, $P = 0.874$), or the 48h
370 test (mean \pm SD = 4.74 \pm 1.97, one-sample t test: $t_{18} = 0.5247$, $P = 0.606$).

371

372 Six robins (3 male, 2 female, 1 sex unknown) improved over time, making at least one less error
373 during the 48 h test as compared to the number of errors made during their initial baseline trial. Five
374 birds (2 male, 2 female, 1 sex unknown) showed no improvement and eight birds (all male)
375 performed worse in the 48h test than in the initial baseline trial.

376

377 Robins did not rely on odour cues to solve the spatial memory task, as the number of errors made by
378 individuals did not differ between the 48 h test (mean \pm SD = 4.74 ± 1.97 , $N = 19$) and the final
379 probe trial (mean \pm SD = 3.89 ± 2.36 , $N = 19$; paired t test: $t_{18} = 1.619$, $P = 0.123$).

380

381 ***Detour Reaching***

382 All 20 robins completed the habituation and training stages (when the cylinder was opaque) in the
383 minimum number of trials required (7 trials). Individuals required between 7 and 59 trials to pass
384 the detour-reaching test when the cylinder was transparent (mean \pm SD = 23.45 ± 15.76 , $N = 20$).

385

386 ***Symbol Discrimination***

387 During the initial preference test, eleven birds chose to remove the lid with the cross symbol first
388 and six birds chose to remove the square. Two of the initial 10 robins that were tested did not pass
389 the symbol task within eight sessions (120 trials). In total, 10 of the 17 robins tested passed the
390 symbol discrimination task. These robins required between 13 and 86 trials to reach the learning
391 criterion of 10 out of 12 consecutive trials correct (mean \pm SD = 55.00 ± 24.29 , $N = 10$). For the 10
392 robins that passed the task, there was a strong correlation between the number of times that they
393 flipped the incorrect lid during their first three sessions (45 test trials) and the number of trials that
394 they required to pass the task (Spearman rank correlation $r_s = 0.832$, $N = 10$, $P = 0.003$). This
395 relationship also held when the two birds that completed 120 trials without passing the task were
396 included in the analysis (Spearman rank correlation $r_s = 0.676$, $N = 12$, $P = 0.016$).

397

398 To enable us to include all 17 robins in subsequent analyses, we used the number of errors made
399 during the first 45 trials of the task as a performance measure. During the initial 45 trials, the
400 number of trials in which robins flipped the incorrect lid ranged between 13 and 25 (mean \pm SD =
401 19.41 ± 3.71 , $N = 17$).

402

403 *Motivation and Neophobia*

404 All robins took a mealworm from the scale both before and after every test session, suggesting that
405 they remained motivated for food rewards throughout all test sessions. For each robin we calculated
406 a mean weight from all weight measurements taken during the experiment. This mean weight
407 ranged between mean \pm SD = 25.5 ± 1.2 g for the lightest bird and mean \pm SD = 30.9 ± 0.7 g for the
408 heaviest (mean \pm SD = 29.0 ± 1.2 g, $N = 19$; one bird never learned to hop slowly enough onto the
409 scale to record his weight). There were no significant correlations between the robins' mean
410 weights and their performances in any of the tasks ($0.216 > P < 0.909$; Bonferroni corrected $\alpha =$
411 0.008 for the six comparisons). We also had tarsus measurements for 11 robins; we found no
412 significant correlations between body condition (defined as the ratio of the body mass to tarsus
413 length; Cauchard et al., 2013) and performances in any of the tasks except colour discrimination
414 (Spearman rank correlation $r_s = -0.770$, $N = 11$, $P = 0.006$; for the other correlations $0.316 > P <$
415 0.935). However, when we removed one bird with a colour preference for red from this analysis
416 (see Fig.2; these birds outperformed all others), the correlation was no longer significant (Spearman
417 rank correlation $r_s = -0.693$, $N = 10$, $P = 0.026$; Bonferroni corrected $\alpha = 0.008$).

418

419 When approaching a novel apparatus for the first time, the quickest robin took on average mean \pm
420 SD = 3.2 ± 1.9 s while the slowest robin took mean \pm SD = 21.4 ± 17.6 s. Heavier robins were on
421 average slower to approach novel objects (Spearman rank correlation $r_s = 0.518$, $N = 19$, $P = 0.023$),
422 but this correlation did not hold for body condition (Spearman rank correlation $r_s = 0.495$, $N = 11$, P
423 = 0.122). The robins' mean latency to approach new apparatuses was not significantly correlated

424 with performance measures for any of the tasks ($0.186 > P < 0.977$; Bonferroni corrected $\alpha = 0.008$
425 for the six comparisons).

426 427 *Relationships between Individual Performances across Cognitive Tasks*

428 Individual performances were positively correlated in the majority of pair-wise comparisons (Table
429 2). However, none of these correlations were significant (Bonferroni-corrected $\alpha = 0.003$). The
430 directional pattern of correlations remained largely consistent when only the performances of the 16
431 robins that completed all tasks were analysed (appendix Table A1).

432 433 *Variance in Cognitive Test Battery Performance*

434 For the PCA of the performances of the 16 robins that participated in all tasks, two components
435 were extracted with eigenvalues >1 . All task performances loaded positively onto the first
436 component, albeit weakly for the motor task and symbol discrimination task (Table 3). This pattern
437 of factor loadings was highly unlikely to occur at random. In 10 000 random simulations, only
438 0.0048% had a larger mean loading on the first component extracted in the PCA (Fig. 3a) and only
439 0.0056% had a smaller standard deviation for the first component loadings (Fig. 3b). The first
440 component extracted from the PCA explained 34.46% of the total variance in the task performances
441 of the 16 robins. Performances in the motor task and symbol discrimination loaded positively on the
442 second component, which captured a further 24.44% of the variance.

443
444 In the spatial memory task there was no clear evidence that robins learned the relevant spatial cue
445 (see Spatial Memory results above). This does not preclude the possibility that robins utilised
446 spatial memory in the task, but it remains unclear whether the task provides an informative
447 cognitive measure. We therefore ran an additional PCA from which we excluded the spatial
448 memory task. The results were consistent with the PCA including all tasks (Table 3).

449

450 In the colour discrimination task, robins that initially preferred the red lid (and hence had the blue
451 lid rewarded in the test) were by far the fastest to solve the task (Fig. 2). This result suggests that
452 rewarding blue lids may have biased these birds' performances in the colour discrimination task.
453 We therefore ran an additional PCA from which we excluded those birds that chose the red lid in
454 the colour preference test ($N = 13$). This produced results that were consistent with the PCA
455 including all 16 individuals (Table 3).

456

457 Finally, motor task performance may have been influenced by an individual's prior experience, as
458 robins typically turn over leaves as they forage for invertebrate prey in leaf litter (Hunt et al., 2008).
459 We therefore conducted a PCA on a more conservative dataset of cognitive performances, which
460 excluded birds with a preference for red, as well as the performance scores from the motor task.
461 Two eigenvalues >1 were extracted and the factor loadings of the tasks were consistent with the
462 other PCA analyses. However, the first component now explained 45.67% of the data (Table 3).

463

464 **DISCUSSION**

465 We investigated individual variation in the cognitive performances of 20 wild North Island robins.
466 Robins voluntarily participated in a test battery comprised of six tasks: a motor task, colour
467 discrimination and reversal learning, symbol discrimination, spatial memory and inhibitory control.
468 Individuals differed greatly in their ability to solve these tasks. For the 16 robins that completed all
469 tasks, we found weak, non-significant positive correlations between most task performances. In
470 human psychometric testing, individuals' test scores are positively correlated across tasks assessing
471 several cognitive domains, with a general factor typically accounting for 40% of total variance
472 (Deary et al., 2006; Plomin, 2001). In the PCA of 16 robins' performances, all tasks loaded
473 positively on the first component extracted with an eigenvalue >1 and this component captured over
474 34% of the total variance in task performance. Our results are consistent with those of previous
475 studies that have been claimed as evidence for g in non-human animals (e.g. Galsworthy et al.,
476 2005b; Isden et al., 2013; Matzel et al., 2003). However, in contrast to these previous non-human

477 studies, we tested our PCA results against 10 000 random simulations and demonstrated that all
478 tasks uniformly and positively loading on the first component extracted is highly unlikely to be a
479 random occurrence. Thus we provide the first evidence for *g* in a food-hoarding species.

480

481 It has been suggested that *g* loadings are higher for those tasks that are more “cognitively complex”
482 (Plomin, 2001). Interestingly, the task with the lowest loading on the first component extracted was
483 the motor task in our test battery. A low *g* loading is consistent with the suggestion that the motor
484 task is unlikely to be a good measure of cognitive ability for an insectivorous litter foraging species
485 (Boogert, Anderson, et al., 2011), as prior foraging experience may have influenced robins’ learning
486 performance in this task. Indeed, removing the performance scores for the motor task from the PCA
487 increased the amount of variance explained by the first component extracted to almost 46%.
488 Additionally, the lack of positive correlations with other task performances could be further
489 evidence that motor task performance is strongly influenced by prior experience in wild birds
490 (Boogert, Anderson, et al., 2011).

491

492 Our finding that robins presented with rewarded blue lids in the colour discrimination task
493 outperformed all other birds validates existing concerns regarding the use of colour cues in
494 cognitive tests of animals whose prior experience of particular colours is unknown (see Rowe and
495 Healy, 2014 for a detailed critique). Blue is not a commonly occurring colour in New Zealand forest
496 flora and fauna, while red, green and yellow are much more prevalent. Blue might thus be more
497 salient in the natural setting where we carried out our tests and robins have neophilic tendencies
498 which may attract them to this novel colour cue. For any future colour discrimination tasks on
499 robins we would advocate the use of grey scale cues, as this may help to reduce such confounding
500 effects on learning performance (Rowe & Healy, 2014).

501

502 Reversal learning tasks require an animal to stop responding to a previously rewarded stimulus and
503 switch its responses to a previously *unrewarded* stimulus (Bond et al., 2007). By contrast, in

504 discrimination tasks there is no (known) previous history of reinforcement that the animal must
505 overcome to learn the reward contingencies. It is therefore unsurprising that the robins required
506 more trials to learn the colour reversal than they did for the original colour discrimination task. This
507 pattern has also been found for song sparrows (Boogert, Anderson, et al., 2011), corvids (Bond et
508 al., 2007), Darwin's finches (Tebbich et al., 2010) and spotted bowerbirds (Isden et al., 2013).
509 Reversal learning has been argued to provide a measure of inhibitory control (Bond et al., 2007), as
510 animals are required to inhibit a previously rewarded response to learn the reversed contingency.
511 Indeed, our results provide evidence that inhibitory control may be an important component of
512 reversal learning in robins, as the robins' performance in the reversal task was positively correlated
513 with performance in the detour-reaching task (although not significantly so).

514

515 The symbol discrimination took longer for the robins to learn than the colour discrimination. The
516 symbol stimuli that we used were matched in all aspects except the arrangement of the lines. Thus
517 the birds may have been slower to learn the reward contingency because there were perceptual or
518 attentional constraints related to the symbol design (e.g. some robins may not have perceived the
519 symbols as being markedly different, or the differences were insufficiently salient to capture robins'
520 attention). However, Isden and colleagues (2013) used block symbols (rather than lines) and also
521 found that spotted bowerbirds took longer to learn the symbol discrimination compared to their
522 learning speed in a colour discrimination task. Additionally, in a previous study of avoidance
523 learning in chicks (*Gallus gallus domesticus*), birds attended to the colouration, rather than the
524 black pattern, of an aposematic stimulus that contained both colour and pattern (Aronsson &
525 Gamberale-Stille, 2008), while a study of "same-different" conditional discrimination in pigeons
526 found that birds performed worse in transfer tests that used shape stimuli as opposed to those that
527 used colour stimuli (Cook, Cavoto, & Cavoto, 1995). Taken together with our own results, these
528 studies suggest that avian symbol discrimination task performances may be strongly affected by
529 factors such as perception and attention. It is particularly interesting to note that the robins' symbol
530 task performance in our study was loaded on the second component extracted. This is potentially

531 further evidence that symbol discrimination performance was influenced by cognitive factors
532 different from those that affected performance on the four tasks that loaded heavily on the first
533 principal component extracted. These additional cognitive processes may have been attentional,
534 perceptual, or a combination of both.

535

536 Robin food-hoarding behaviour peaks outside of the breeding season (Steer & Burns, 2008). As we
537 ran our test battery outside of the breeding season, our expectation was that robins would perform
538 well at memorising spatial locations. Six robins made fewer errors over time in the spatial memory
539 test, suggesting they may have memorized the spatial location. **However, at the group level there**
540 **was no evidence that the robins' search behaviour deviated from a random search expectation.** In a
541 previous study of avian spatial memory, song sparrows showed a reduction in the number of errors
542 made over time as they searched for a single reward in one of six possible locations over eight
543 consecutive days, suggesting that they had learned the spatial cue (Sewall et al., 2013). It is unclear
544 whether our protocol, which ran for only three days, allowed sufficient trials for the robins to
545 memorize the spatial location of the reward. Additionally, the size of the spatial grid may have been
546 a factor in the robins' poor performance in this task, as robins typically cache in branch-trunk axils
547 and previous experiments have shown that these caches are spaced approximately 6 m apart on
548 average (Van Horik & Burns, 2007). Thus the foraging grid may have been at a scale that was
549 inappropriate if robins' spatial memory is optimised to recall widely spaced cache sites. For future
550 studies of robin spatial memory we would advocate running the task over additional days and
551 increasing the distance between spatial locations.

552

553 A previous study of food-storing birds found that spatial memory, which is reliant on hippocampal
554 function, was impaired by developmental nutritional deficits, while associative learning for colour
555 cues was not, suggesting that these are distinct cognitive processes (Pravosudov et al., 2005). By
556 contrast, in our study we found that spatial memory performance loaded on the same component as
557 colour discrimination performance, rather than a different component as might be expected if the

558 cognitive processes required for these tasks differed. It is possible that our protocol did not enable
559 robins to utilise spatial memory for the reasons described above. Conceivably robins may have
560 relied solely on very subtle visual cues to solve the task, which would explain why their
561 performance on this task did not load on a separate principal component in our analyses. This issue
562 thus requires further study.

563

564 Motivational factors have the potential to greatly influence individual performance in a cognitive
565 task (Rowe & Healy, 2014). This is particularly problematic in the wild, where it is exceedingly
566 difficult to standardise levels of motivation between individuals participating in cognitive
567 experiments (Boogert et al., 2010; Isden et al., 2013). The authors of the bowerbird study suggested
568 that lack of motivation may have played a role in the poor performance of males in some tasks
569 (Isden et al., 2013). However, in our study motivational differences may account for less of the
570 inter-individual variation in task performances. As a food-hoarding species (Steer & Burns, 2008;
571 Van Horik & Burns, 2007), robins may remain consistently motivated for food rewards even when
572 satiated. Indeed, many robins switched from eating to storing food for later consumption during test
573 sessions. Moreover, all robins always took food during our motivation checks at the end of each test
574 session. However, we cannot conclude from this that other non-cognitive factors, such as previous
575 experience or hormonal status, had little or no influence on a robin's task performances.

576

577 In common with previous avian test batteries (Boogert, Anderson, et al., 2011; Isden et al., 2013),
578 our study included tests of various types of learning, as well as inhibitory control. In future test
579 batteries it would be worthwhile testing additional cognitive domains (e.g. social cognition) and
580 incorporating additional types of task. A recent meta-analysis of test battery data from captive apes
581 found no evidence of a general intelligence factor (Herrmann & Call, 2012). Instead, tasks requiring
582 inferential reasoning clustered together, while those that were learning-based formed a distinct
583 cluster (Herrmann & Call, 2012). Tasks requiring inference have previously been implemented for
584 captive food-hoarding birds (e.g. Shaw et al., 2013). Adapting such protocols for use with wild

585 birds would allow for the inclusion of inference-based tasks in future test batteries, thereby adding
586 to our current understanding of the structure of avian intelligence. The development of an avian
587 cognitive test battery is still in its early stages and our study builds on pioneering work (Boogert,
588 Anderson, et al., 2011; Isden et al., 2013). As we have highlighted, there is still ample scope for
589 fine-tuning field-based avian cognitive test methodologies. Nonetheless, our results provide some of
590 the first evidence that a factor analogous to human *g* may underpin cognitive performance in a food-
591 hoarding bird.

592

593 ACKNOWLEDGEMENTS

594 This research was funded by a Fast-Start grant from the Marsden Fund of the Royal Society of New
595 Zealand (VUW1304) and a Rutherford Foundation New Zealand Postdoctoral Fellowship to RCS.
596 We thank Raewyn Empson and the Karori Wildlife Sanctuary Trust for permission to work at
597 Zealandia Wildlife Sanctuary. We thank Annette Harvey and Matu Booth for advice and assistance
598 with banding robins, Neville Higgison and Shaun Graham for assistance with apparatus
599 construction and Michael Corballis for discussion and statistical advice.

600

601 REFERENCES

- 602 Amici, F., Barney, B., Johnson, V. E., Call, J., & Aureli, F. (2012). A modular mind? A test using
603 individual data from seven primate species. *PloS One*, 7(12), e51918.
604 doi:10.1371/journal.pone.0051918
- 605 Aronsson, M., & Gamberale-Stille, G. (2008). Domestic chicks primarily attend to colour, not
606 pattern, when learning an aposematic coloration. *Animal Behaviour*, 75(2), 417–423.
607 doi:10.1016/j.anbehav.2007.05.006
- 608 Banerjee, K., Chabris, C. F., Johnson, V. E., Lee, J. J., Tsao, F., & Hauser, M. D. (2009). General
609 intelligence in another primate: individual differences across cognitive task performance in a
610 New World monkey (*Saguinus oedipus*). *PloS One*, 4(6), e5883.
611 doi:10.1371/journal.pone.0005883
- 612 Bond, A. B., Kamil, A. C., & Balda, R. P. (2007). Serial reversal learning and the evolution of
613 behavioral flexibility in three species of North American corvids (*Gymnorhinus*
614 *cyanocephalus*, *Nucifraga columbiana*, *Aphelocoma californica*). *Journal of Comparative*
615 *Psychology*, 121(4), 372–379. doi:10.1037/0735-7036.121.4.372

- 616 Boogert, N. J., Anderson, R. C., Peters, S., Searcy, W. A., & Nowicki, S. (2011). Song repertoire
617 size in male song sparrows correlates with detour reaching, but not with other cognitive
618 measures. *Animal Behaviour*, *81*(6), 1209–1216. doi:10.1016/j.anbehav.2011.03.004
- 619 Boogert, N. J., Fawcett, T. W., & Lefebvre, L. (2011). Mate choice for cognitive traits: a review of
620 the evidence in nonhuman vertebrates. *Behavioral Ecology*, *22*(3), 447–459.
621 doi:10.1093/beheco/arq173
- 622 Boogert, N. J., Giraldeau, L.-A., & Lefebvre, L. (2008). Song complexity correlates with learning
623 ability in zebra finch males. *Animal Behaviour*, *76*(5), 1735–1741.
624 doi:10.1016/j.anbehav.2008.08.009
- 625 Boogert, N. J., Monceau, K., & Lefebvre, L. (2010). A field test of behavioural flexibility in
626 Zenaida doves (*Zenaida aurita*). *Behavioural Processes*, *85*(2), 135–141.
627 doi:10.1016/j.beproc.2010.06.020
- 628 Bouchard, J., Goodyer, W., & Lefebvre, L. (2007). Social learning and innovation are positively
629 correlated in pigeons (*Columba livia*). *Animal Cognition*, *10*(2), 259–66. doi:10.1007/s10071-
630 006-0064-1
- 631 Burns, K. C., & Steer, J. (2006). Dominance rank influences food hoarding in New Zealand Robins
632 *Petroica australis*. *Ibis*, *148*(2), 266–272. doi:10.1111/j.1474-919X.2006.00529.x
- 633 Cauchard, L., Boogert, N. J., Lefebvre, L., Dubois, F., & Doligez, B. (2013). Problem-solving
634 performance is correlated with reproductive success in a wild bird population. *Animal*
635 *Behaviour*, *85*(1), 19–26. doi:10.1016/j.anbehav.2012.10.005
- 636 Chandra, S. B., Hosler, J. S., & Smith, B. H. (2000). Heritable variation for latent inhibition and its
637 correlation with reversal learning in honeybees (*Apis mellifera*). *Journal of Comparative*
638 *Psychology*, *114*, 86–97.
- 639 Cheke, L. G., Bird, C. D., & Clayton, N. S. (2011). Tool-use and instrumental learning in the
640 Eurasian jay (*Garrulus glandarius*). *Animal Cognition*, *14*, 441–455. doi:10.1007/s10071-011-
641 0379-4
- 642 Cheke, L. G., & Clayton, N. S. (2012). Eurasian jays (*Garrulus glandarius*) overcome their current
643 desires to anticipate two distinct future needs and plan for them appropriately. *Biology Letters*,
644 *8*(2), 171–175. doi:10.1098/rsbl.2011.0909
- 645 Cole, E. F., Morand-Ferron, J., Hinks, A. E., & Quinn, J. L. (2012). Cognitive ability influences
646 reproductive life history variation in the wild. *Current Biology*, *22*(19), 1808–1812.
647 doi:10.1016/j.cub.2012.07.051
- 648 Cook, R., Cavoto, K., & Cavoto, B. (1995). Same-different texture discrimination and concept
649 learning by pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, *21*,
650 253–260. Retrieved from <http://psycnet.apa.org/journals/xan/21/3/253/>
- 651 Deary, I. J., Penke, L., & Johnson, W. (2010). The neuroscience of human intelligence differences.
652 *Nature Reviews. Neuroscience*, *11*(3), 201–11. doi:10.1038/nrn2793
- 653 Deary, I. J., Spinath, F. M., & Bates, T. C. (2006). Genetics of intelligence. *European Journal of*
654 *Human Genetics*, *14*(6), 690–700. doi:10.1038/sj.ejhg.5201588

- 655 Galsworthy, M. J., Paya-Cano, J. L., Liu, L., Monleón, S., Gregoryan, G., Fernandes, C., ...
656 Plomin, R. (2005). Assessing reliability, heritability and general cognitive ability in a battery
657 of cognitive tasks for laboratory mice. *Behavior Genetics*, 35(5), 675–692.
658 doi:10.1007/s10519-005-3423-9
- 659 Garland, A., Low, J., Armstrong, N., & Burns, K. C. (2014). Wild robins (*Petroica longipes*)
660 respond to human gaze. *Animal Cognition*, 17(5), 1149–1156. doi:10.1007/s10071-014-0747-y
- 661 Garland, A., Low, J., & Burns, K. C. (2012). Large quantity discrimination by North Island robins
662 (*Petroica longipes*). *Animal Cognition*, 15, 1129–1140. doi:10.1007/s10071-012-0537-3
- 663 Hauser, M. D. (1999). Perseveration, inhibition and the prefrontal cortex: a new look. *Current*
664 *Opinion in Neurobiology*, 9(2), 214–222. doi:10.1016/S0959-4388(99)80030-0
- 665 Herrmann, E., & Call, J. (2012). Are there geniuses among the apes? *Philosophical Transactions of*
666 *the Royal Society B: Biological Sciences*, 367, 2753–2761. doi:10.1098/rstb.2012.0191
- 667 Herrmann, E., Call, J., Hernández-Lloreda, M. V., Hare, B., & Tomasello, M. (2007). Humans have
668 evolved specialized skills of social cognition: the cultural intelligence hypothesis. *Science*,
669 317(5843), 1360–1366. doi:10.1126/science.1146282
- 670 Hunt, S., Low, J., & Burns, K. C. (2008). Adaptive numerical competency in a food-hoarding
671 songbird. *Proceedings of the Royal Society B: Biological Sciences*, 275(1649), 2373–2379.
672 doi:10.1098/rspb.2008.0702
- 673 Isden, J., Panayi, C., Dingle, C., & Madden, J. (2013). Performance in cognitive and problem-
674 solving tasks in male spotted bowerbirds does not correlate with mating success. *Animal*
675 *Behaviour*, 86(4), 829–838. doi:10.1016/j.anbehav.2013.07.024
- 676 Keagy, J., Savard, J.-F., & Borgia, G. (2009). Male satin bowerbird problem-solving ability predicts
677 mating success. *Animal Behaviour*, 78(4), 809–817. doi:10.1016/j.anbehav.2009.07.011
- 678 Keagy, J., Savard, J.-F., & Borgia, G. (2011). Complex relationship between multiple measures of
679 cognitive ability and male mating success in satin bowerbirds, *Ptilonorhynchus violaceus*.
680 *Animal Behaviour*, 81(5), 1063–1070. doi:10.1016/j.anbehav.2011.02.018
- 681 Kembel, S., Ackerly, D., Blomberg, S., Cornwell, W., Cowan, P., Helmus, M., ... Webb, C. (2014).
682 picante: R tools for integrating phylogenies and ecology. Retrieved from
683 <http://www.proesa.org.co/CRAN/web/packages/picante/picante.pdf>
- 684 Locurto, C., Fortin, E., & Sullivan, R. (2003). The structure of individual differences in
685 Heterogeneous Stock mice across problem types and motivational systems. *Genes, Brain and*
686 *Behavior*, 2(1), 40–55. doi:10.1034/j.1601-183X.2003.00006.x
- 687 MacLean, E. L., Hare, B., Nunn, C. L., Addessi, E., Amici, F., Anderson, R. C., ... Zhao, Y. (2014).
688 The evolution of self-control. *Proceedings of the National Academy of Sciences of the United*
689 *States of America*, 111(20), E2140–8. doi:10.1073/pnas.1323533111
- 690 Maloney, R. F., & Mclean, I. G. (1995). Historical and experimental learned predator recognition in
691 free-living New Zealand robins. *Animal Behaviour*, 50, 1193–1201.
- 692 Matzel, L. D., Han, Y. R., Grossman, H., Karnik, M. S., Patel, D., Scott, N., ... Gandhi, C. C.
693 (2003). Individual differences in the expression of a “general” learning ability in mice. *The*

- 694 *Journal of Neuroscience*, 23(16), 6423–6433. Retrieved from
695 <http://www.ncbi.nlm.nih.gov/pubmed/12878682>
- 696 McGavin, S. (2009). Density and pair fidelity in a translocated population of North Island robin
697 (*Petroica longipes*). *Notornis*, 56, 206–212.
- 698 Ostojić, L., Shaw, R. C., Cheke, L. G., & Clayton, N. S. (2013). Evidence suggesting that desire-
699 state attribution may govern food sharing in Eurasian jays. *Proceedings of the National*
700 *Academy of Sciences of the United States of America*, 110, 4123–4128.
701 doi:10.1073/pnas.1209926110
- 702 Plomin, R. (2001). The genetics of g in human and mouse. *Nature Reviews Neuroscience*,
703 2(February), 136–141. Retrieved from
704 http://www.nature.com/nrn/journal/v2/n2/abs/nrn0201_136a.html
- 705 Plomin, R., & Spinath, F. M. (2002). Genetics and general cognitive ability (g). *Trends in Cognitive*
706 *Sciences*, 6(4), 169–176. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/11912040>
- 707 Pravosudov, V. V., Lavenex, P., & Omanska, A. (2005). Nutritional deficits during early
708 development affect hippocampal structure and spatial memory later in life. *Behavioral*
709 *Neuroscience*, 119(5), 1368–1374. doi:10.1037/0735-7044.119.5.1368
- 710 Rowe, C., & Healy, S. D. (2014). Measuring variation in cognition. *Behavioral Ecology*, 25, 1287–
711 1292. doi:10.1093/beheco/aru090
- 712 Sanford, K., & Clayton, N. S. (2008). Motivation and memory in zebra finch (*Taeniopygia guttata*)
713 foraging behavior. *Animal Cognition*, 11(2), 189–198. doi:10.1007/s10071-007-0106-3
- 714 Schmitt, V., Pankau, B., & Fischer, J. (2012). Old world monkeys compare to apes in the primate
715 cognition test battery. *PloS One*, 7(4), e32024. doi:10.1371/journal.pone.0032024
- 716 Sewall, K. B., Soha, J., Peters, S., & Nowicki, S. (2013). Potential trade-off between vocal
717 ornamentation and spatial ability in a songbird. *Biology Letters*, 9, 20130344. Retrieved from
718 <http://rsbl.royalsocietypublishing.org/content/9/4/20130344.short>
- 719 Shaw, R. C., & Clayton, N. S. (2013). Careful cachers and prying pilferers: Eurasian jays (*Garrulus*
720 *glandarius*) limit auditory information available to competitors. *Proceedings of the Royal*
721 *Society B: Biological Sciences*, 280, 20122238. doi:<http://dx.doi.org/10.1098/rspb.2012.2238>
- 722 Shaw, R. C., Plotnik, J. M., & Clayton, N. S. (2013). Exclusion in corvids: the performance of food-
723 caching Eurasian jays (*Garrulus glandarius*). *Journal of Comparative Psychology*, 127(4),
724 428–435. doi:10.1037/a0032010
- 725 Shettleworth, S. J. (2010). *Cognition, Evolution, and Behaviour* (Second.). Oxford: Oxford
726 University Press.
- 727 Smulders, T. V., Gould, K. L., & Leaver, L. A. (2010). Using ecology to guide the study of
728 cognitive and neural mechanisms of different aspects of spatial memory in food-hoarding
729 animals. *Philosophical Transactions of the Royal Society of London. Series B, Biological*
730 *Sciences*, 365(1542), 883–900. doi:10.1098/rstb.2009.0211

- 731 Steer, J., & Burns, K. C. (2008). Seasonal variation in male-female competition, cooperation and
732 selfish hoarding in a monogamous songbird. *Behavioral Ecology and Sociobiology*, *62*(7),
733 1175–1183. doi:10.1007/s00265-008-0546-z
- 734 Tebbich, S., Sterelny, K., & Teschke, I. (2010). The tale of the finch: adaptive radiation and
735 behavioural flexibility. *Philosophical Transactions of the Royal Society of London. Series B,*
736 *Biological Sciences*, *365*(1543), 1099–1109. doi:10.1098/rstb.2009.0291
- 737 Thornton, A. (2014). How and why are some species so smart? A comment on Rowe and Healy.
738 *Behavioral Ecology*, *00*(May), 1–2. doi:10.1093/beheco/aru115
- 739 Thornton, A., Isden, J., & Madden, J. R. (2014). Toward wild psychometrics: linking individual
740 cognitive differences to fitness. *Behavioral Ecology*, *25*, 1299–1301.
741 doi:10.1093/beheco/aru095
- 742 Thornton, A., & Lukas, D. (2012). Individual variation in cognitive performance: developmental
743 and evolutionary perspectives. *Philosophical Transactions of the Royal Society of London.*
744 *Series B, Biological Sciences*, *367*(1603), 2773–2783. doi:10.1098/rstb.2012.0214
- 745 Tillé, Y., Newman, J. a., & Healy, S. D. (1996). New tests for departures from random behavior in
746 spatial memory experiments. *Animal Learning & Behavior*, *24*(3), 327–340.
747 doi:10.3758/BF03198981
- 748 Van Horik, J., & Burns, K. C. (2007). Cache spacing patterns and reciprocal cache theft in New
749 Zealand robins. *Animal Behaviour*, *73*(6), 1043–1049. doi:10.1016/j.anbehav.2006.07.014
- 750
- 751
- 752

753 **Figure 1** The test apparatuses, with a robin for scale: (a) the foraging grid and white lids used for
754 the motor task, (b) the grid and coloured lids used for the colour discrimination and reversal tasks,
755 (c) the two grids and lids used for the spatial memory task, (d) the opaque cylinder used in the
756 training phase of the detour reach task, (e) the transparent cylinder used in the detour reach test
757 phase (f) and the scales used to check motivation

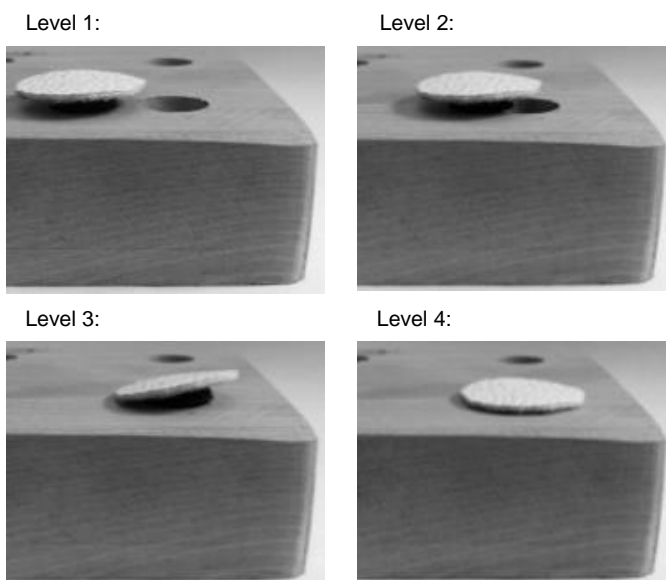
758

759 **Figure 2** The number of trials that robins required to solve the colour discrimination task, grouped
760 by their initial lid colour preference (red: $N = 4$; blue: $N = 8$; yellow: $N = 4$; green: $N = 4$; *whiskers*
761 range, *boxes* upper quartile, median and lower quartile)

762

763 **Figure 3** Histograms of (a) the mean factor loadings and (b) the standard deviation of the factor
764 loadings for the first unrotated component extracted in the 10 000 random PCA simulations. The
765 observed (a) mean and (b) standard deviation for the first component loadings for our data are
766 indicated by the arrows

767 **Figure A1** The position of the lid relative to the well in levels 1-4 of the novel motor task



768

769 **Table 1** The four stages of the motor task that were used to train robins to flip the lids off the
 770 foraging grid (following Boogert et al. 2008). For images matching the descriptions of the lid
 771 positions, refer to figure A1
 772

level	description	pass criteria
1	lids positioned next to the wells	3-4 mealworms retrieved in 3 consecutive trials
2	half of each well covered by a lid	3-4 mealworms retrieved in 3 consecutive trials
3	mealworms hidden from view by lids propped up on one side	3-4 mealworms retrieved in 3 consecutive trials
4	lids' vinyl bumpers fitted into the wells	3-4 mealworms retrieved in 6 out of 7 consecutive trials

773
 774
 775
 776

Table 2 Spearman rank correlation matrix of all cognitive task performances.

		motor task	colour discrimination	colour reversal	symbol discrimination	spatial memory
colour discrimination	r_s	0.016				
	P	0.948				
	N	20				
colour reversal	r_s	-0.032	0.294			
	P	0.895	0.208			
	N	20	20			
symbol discrimination	r_s	0.412	0.124	0.130		
	P	0.101	0.635	0.618		
	N	17	17	17		
spatial memory	r_s	-0.178	0.273	0.314	-0.034	
	P	0.465	0.258	0.190	0.902	
	N	19	19	19	16	
detour reaching	r_s	-0.008	0.253	0.380	0.106	0.314
	P	0.972	0.282	0.099	0.685	0.190
	N	20	20	20	17	19

777
 778

779 **Table 3** Results of the unrotated principal components analysis for the robins that completed all
 780 tasks (N = 16), for all robins but excluding the spatial memory task (N = 16), for all tasks but
 781 excluding the robins with a red preference in the colour preference test (N = 13) and excluding the
 782 motor task performances and robins with a preference for red (N = 13). The loadings and
 783 percentage of variance explained for each Principal Component (PC) with an eigenvalue > 1 are
 784 shown. Loadings > 0.6 are in bold

task	all tasks		excluding the spatial memory task		excluding robins with red preference		excluding motor task and red preference	
	PC1	PC2	PC1	PC2	PC1	PC2	PC1	PC2
motor task	0.231	0.887	0.435	0.804	0.199	0.879	-	-
colour discrimination	0.660	0.084	0.632	-0.082	0.735	0.367	0.690	0.427
colour reversal	0.631	-0.274	0.569	-0.487	0.682	-0.344	0.713	-0.309
symbol discrimination	0.411	0.673	0.648	0.471	0.426	0.590	0.362	0.795
spatial memory	0.727	-0.184	-	-	0.749	-0.139	0.759	-0.147
detour reach	0.695	-0.333	0.612	-0.556	0.717	-0.497	0.770	-0.325
Eigenvalue	2.067	1.466	1.723	1.421	2.302	1.641	2.284	1.038
% of variance explained	34.46	24.44	34.47	28.42	38.36	27.35	45.69	20.77

785 **Table A1** Spearman rank correlation matrices of cognitive task performances for the 16 robins that
 786 completed all tasks
 787

		motor task	colour discrimination	colour reversal	symbol discrimination	spatial memory
colour discrimination	r_s	0.237				
	P	0.377				
	N	16				
colour reversal	r_s	-0.073	0.396			
	P	0.788	0.129			
	N	16	16			
symbol discrimination	r_s	0.502	0.059	0.165		
	P	0.048	0.829	0.541		
	N	16	16	16		
spatial memory	r_s	0.064	0.371	0.456	-0.034	
	P	0.812	0.157	0.076	0.902	
	N	16	16	16	16	
detour reaching	r_s	-0.089	0.302	0.411	0.079	0.462
	P	0.744	0.255	0.114	0.771	0.072
	N	16	16	16	16	16

788
789