

1 Inhibitory control performance is repeatable across years and 2 contexts in a wild bird population

3 Authors: Gabrielle L. Davidson^{a,b,*}, Michael S. Reichert^{a, e}, Jennifer R. Coomes^a, Ipek G.
4 Kulahci^{a,d}, Iván de la Hera^{a,e}, John L. Quinn^{a,e*}

5 * Correspondence: Gabrielle Davidson email: gd339@cam.ac.uk; John Quinn email: j.quinn@ucc.ie.

6 Tel: +44(0)1223 333550

7 a. School of Biological, Earth and Environmental Sciences, Distillery Fields, North Mall,
8 University College Cork, Cork, Ireland.

9 b. Department of Psychology, Downing Street, University of Cambridge, Cambridge, UK.

10 c. Department of Integrative Biology, Oklahoma State University, USA

11 d. University of Notre Dame, USA

12 e. Environmental Research Institute, University College Cork, Cork, Ireland

13

14 Keywords: great tits, inhibitory control, cognition, detour task, repeatability, wild

15 **Abstract**

16 Inhibitory control is one of several cognitive mechanisms required for self-regulation, decision
17 making and attention towards tasks. Linked to a variety of maladaptive behaviours in humans,
18 inhibitory control is expected to influence behavioural plasticity in animals in the context of foraging,
19 social interaction, or responses to sudden changes in the environment. One widely used cognitive
20 assay, the 'detour task', putatively tests inhibitory control. In this task, subjects must avoid
21 impulsively touching transparent barriers positioned in front of food, and instead access the food by
22 an alternative but known route. Recently it has been suggested that the detour task is unreliable and
23 measures factors unrelated to inhibitory control, including motivation, previous experience and
24 persistence. Consequently, there is growing uncertainty as to whether this task leads to erroneous
25 interpretations about animal cognition and its links with socio-ecological traits. To address these
26 outstanding concerns, we designed a variant of the detour task for wild great tits (*Parus major*) and
27 deployed it at the nesting site of the same individuals across two spring seasons. This approach
28 eliminated the use of food rewards, limited social confounds, and maximised motivation. We
29 compared task performance in the wild with their performance in captivity when tested using the
30 classical cylinder detour task during the non-breeding season. Task performance was temporally and
31 contextually repeatable, and none of the confounds had any significant effect on performance, nor

32 did they drive any of the observed repeatable differences among individuals. These results support
33 the hypothesis that our assays captured intrinsic differences in inhibitory control. Instead of
34 ‘throwing the detour task out with the bathwater’, we suggest confounds are likely system and
35 experimental-design specific, and that assays for this potentially fundamental but largely overlooked
36 source of behavioural plasticity in animal populations, should be validated and refined for each study
37 system.

38 **Introduction**

39 Inhibitory control is a well-known form of self-regulation that affects decision making and is linked to
40 a variety of maladaptive behaviours in humans, including addiction (Jentsch & Taylor, 1999; Loeber
41 & Duka, 2009), obesity (Epstein et al., 2008; Houben et al., 2014) and asocial emotional responses
42 (Tang & Schmeichel, 2014). Inhibitory control is increasingly being used to study animal behaviour
43 due to its potential role in socio-ecological processes (reviewed in Kabadayi et al., 2018), including
44 foraging plasticity (Coomes et al., 2020), dietary breadth (MacLean et al., 2014; Van Horik et al.,
45 2018), and social interactions (Amici et al., 2013, 2018; Reddy et al., 2015). Moreover, inhibitory
46 control is likely responsive to selection given its links with brain size in primates, and its heritability
47 in humans and birds (Friedman et al., 2008; Langley et al., 2020). However, the inferential power of
48 these studies is dependent on cognitive assays that accurately characterise inhibitory control and
49 whether these assays are applicable to how animals behave in natural settings. Recently, researchers
50 have taken a variety of classical cognitive tasks to the field (e.g. Johnson-Ulrich et al., 2020; Morand-
51 Ferron et al., 2011, 2015, 2016; Muth et al., 2018; Reichert et al., 2020; Sonnenberg et al., 2019;
52 Toledo et al., 2020) where cognition can be assayed under natural conditions with high ecological
53 validity, and welfare concerns and administrative burdens associated with housing animals in
54 captivity are reduced. However, inhibitory control assays in the wild are scant, and to date rely on
55 experimenters in close proximity to habituated animals (Ashton et al., 2018; Shaw et al., 2015).
56 Moreover, there remains uncertainty regarding what is being measured in inhibitory control tasks in
57 animals. For example, variation in performance on one assay of inhibitory control does not always
58 predict performance in another. Possible reasons for this include that inhibitory control is a
59 composite trait, and variations of the task can measure different components of inhibitory control
60 (Bari & Robbins, 2013; Völter et al., 2018); that inhibitory control is tightly linked with many other
61 decision-making processes, such as attention and task switching (Bari & Robbins, 2013); and because
62 non-cognitive processes (e.g. previous experience) confound task performance (Jelbert et al., 2016;
63 van Horik et al., 2020; Van Horik et al., 2018). More generally, the extent to which extraneous
64 variables contribute to or confound cognitive performance may affect captive-bred, temporarily-

65 captive wild individuals, and free-living animals differentially (Morand-Ferron et al., 2016), but
66 comparisons across such experimental contexts have rarely been examined empirically for any
67 cognitive trait (see Benson-Amram et al., 2013; Cauchoix et al., 2017; Forss et al., 2015; McCune et
68 al., 2019; Morand-Ferron et al., 2011; Mouchet & Dingemanse, 2021 for some examples).

69 One way of measuring inhibitory control is the widely used ‘detour’ task. In this task, subjects must
70 avoid and move around a transparent barrier to retrieve a food reward that is positioned directly in
71 front of them, but behind the barrier (Diamond, 1990). A central premise of this task is that the
72 visible reward generates a strong prepotent impulse to approach it directly. This impulse must be
73 inhibited, and the subject must instead move in a direction away from the reward in order to detour
74 around the barrier successfully. A practical advantage of the detour task is that it is less laborious
75 than other cognitive assays, because animals quickly pass the habituation and training phases, and
76 the test phases can be achieved within 10 trials (MacLean et al., 2014) or even three trials (Van Horik
77 et al., 2018). Therefore the detour task has the potential to be a valuable field paradigm for
78 addressing key questions regarding the causes and consequences of individual differences in
79 cognition, provided it is a reliable measure of inhibitory control (Shaw & Schmelz, 2017; Thornton et
80 al., 2014). However, because the classic detour task is dependent on an experimenter re-baiting the
81 apparatus with a food reward, its scope for field-based assays has thus far been restricted to wild,
82 human-habituated New Zealand robins and Australian magpies (e.g. Ashton et al., 2018; Shaw,
83 2017b). Another limitation for cognitive tasks in the field is that invariably these tasks involve
84 providing food rewards where many species visit in groups. Consequently, individual performance is
85 likely subject to social interference, for example through social learning or competition from other
86 group members (Cole & Quinn, 2012; Morand-Ferron & Quinn, 2011; Reichert et al., 2020). To
87 address social interference, individuals can be temporarily brought into captivity, where they are
88 housed in relative isolation. While this approach may solve confounds linked to sociality, they could
89 equally generate different confounding effects on performance, for example through stress and
90 motivation (Butler et al., 2006).

91 Cognition is challenging to assay because the expression of cognitive capabilities is potentially always
92 confounded by some other factor, such as motivation, personality and persistence (Morand-Ferron
93 et al., 2016). This is likely also true for inhibitory control, which plays a role in executive functioning
94 and decision making, works in tandem with additional executive functions such as attention towards
95 the environment (Bari & Robbins, 2013), and belongs to a domain-general set of brain networks
96 involving multiple interacting cognitive processes (Hampshire & Sharp, 2015). The extent to which
97 cognitive tasks that target inhibitory control tap into additional cognitive and non-cognitive

98 processes may therefore be dependent on the type of task used (e.g. Völter et al., 2018), the context
99 (e.g. captivity vs the wild), and temporal aspects (e.g. season). One approach used for quantifying
100 putative cognitive traits generally has been to control for confounds experimentally (e.g. through
101 food deprivation and training), or to independently measure multiple behavioural traits and test
102 their effect on cognitive performance statistically (e.g. Cooke et al., 2021; Reichert et al., 2020). In
103 wild populations, pinpointing confounds is a critical step towards accurately assaying inhibitory
104 control, and interpreting links with functional behaviour, life-history, and evolution.

105 Examining the consistency of task performance has the potential to shed light on task validity.
106 Typically consistency is quantified by estimating repeatability, that is, by estimating the proportion
107 of variance in a trait measured multiple times that explains between-individual, rather than within-
108 individual, variation within the population. Significant repeatability suggests intrinsic among
109 individual differences in the behaviour, and although it sets the upper limit of heritability (Dohm,
110 2002), equally those individual differences could be caused by permanent environment effects
111 (Wilson, 2018). Repeatability may not necessarily be strong proof of task validity since the task could
112 be consistently measuring the same confounds, especially when using the same task under the same
113 conditions over time. For example, problem solving performance was repeatable in great tits but this
114 repeatability was largely explained by experimentally manipulated motivational effects that were
115 present in both repeat measures (Cooke et al., 2021). This is less likely to be a problem, however,
116 when repeat measures are taken under very different conditions (contextual repeatability) that are
117 unlikely to have shared environmental confounds, and especially when intrinsic confounds are also
118 likely to differ (Nakagawa & Schielzeth, 2010). Finally, as discussed above, confounding variables can
119 explain variation in single measures of cognitive performance, but can also act on the between- and,
120 or the within-individual variance component when dealing with multiple measures from the same
121 individual. It follows that a further important test is to establish whether the consistent between-
122 individual differences specifically are driven by these confounding effects (Cooke et al., 2021;
123 Nakagawa & Schielzeth, 2010). Demonstrating repeatability (temporal) and contextual repeatability
124 in performance on the detour task would provide compelling support for a common cognitive basis
125 of these tasks, especially when controlling for potential confounds.

126 The great tit (*Parus major*) provides a model species in ecology and evolution field studies, and has
127 also been used for exploring the evolutionary ecology of cognitive variation. Great tits breed in nest
128 boxes and can be fitted with Passive Integrative Transponder (PIT) identification tags for remote
129 detection at nest boxes using Radio Frequency Identification (RFID) technology. They readily engage
130 with experimental apparatuses in the wild and in captivity (Aplin et al., 2015; Cauchard et al., 2013;

131 Cole et al., 2012; Morand-Ferron et al., 2011; Troisi et al., 2020), allowing for comparisons of
132 cognitive performance across different settings. We designed and presented a modified version of
133 the detour task at the nest box over two years during the breeding season involving minimal
134 experimental disturbance, social interference from conspecifics, and no need for food rewards. We
135 also ran a classic version of the task by testing wild-caught great tits in captivity during one winter
136 season. This approach allowed us to test for both contextual and temporal repeatability, and to
137 examine statistically whether environmental or state variables caused between individual
138 differences in detour task performance, rather than the hypothesised cognitive mechanisms
139 underpinning inhibitory control. These variables included experience, persistence, motivation, body
140 size, and habitat. Moreover, because personality has been shown to affect how animals engage and
141 perform on cognitive tasks (Dougherty & Guillette, 2018; Guillette et al., 2017; Sih & Del Giudice,
142 2012), we also examined whether ‘exploration behaviour in a novel environment’, a commonly used
143 assay of the fast-slow exploration personality-axis (Dingemanse et al., 2002), explained performance
144 on the captive detour task. If these variables did not explain task performance in general, and the
145 repeatability of performance in particular, this would lend support for the detour task’s utility as a
146 robust measure of inhibitory control, where individual performance is not sensitive to bias from
147 extraneous influences.

148 **Methods**

149 *Field study sites*

150 Our study took place across ten distinct woodland sites in the River Bandon Valley, Co. Cork, Ireland.
151 Five sites were mixed deciduous and five were conifer plantations (Table S1, supplementary
152 materials). Nest boxes, hung at approximately 1.5 metres from the ground, were distributed across
153 these sites at a density of two nest boxes per hectare. Nest boxes were monitored for breeding data
154 (lay dates, incubation period, hatching, clutch size, brood size and number of fledglings). Adults were
155 trapped at the nest box between day 10 and 13 post-hatching to measure biometrics (including
156 weight, wing and tarsus length), and to tag individuals with a PIT tag and a coloured ring with a
157 unique alpha-numerical code for identification. Chicks were ringed and weighed on day 15 post-
158 hatching. Experiments took place between 25th May and 25th June 2017 (hereafter year one), and
159 between 26th May and 15th June 2018 (hereafter year two).

160 *Aviary*

161 Birds were caught from nine of the Bandon Valley sites described above between January and early
162 March 2018, four of which were from mixed deciduous and five were conifer plantations. Birds were

163 transported from the field sites to the aviary in cloth bags within 2 hours of being caught and
164 brought into captivity for approximately 3 weeks before being released at the site of capture. Birds
165 were housed individually in 60L x 50W x 62H cm enclosures with two perches above the floor of the
166 cage. Each box was individually lit with an LED light bulb on a 10L/14D cycle, and equipped with its
167 own ventilation system. A small viewing window (14cm diameter) at the top half of the box, and an
168 access flap at the base (10cm) allowed for experimenters to watch the birds without being seen, and
169 to change experimental apparatuses with minimal disturbance to the birds. Birds could also be
170 observed on a video monitor connected to cameras fixed to the top of each box. Birds were
171 provided with vitamin-fortified water, sunflower seeds and peanuts ad libitum. Mealworms
172 (*Tenebrio molitor* larvae) were provided at 3 time points each day (morning, afternoon, evening).
173 Wax worms (*Galleria mellonella* larvae) were only provided as experimental rewards.

174 *Experimental procedures: Detour task in the wild*

175 To measure individuals' detour task performance in the wild, we designed a detour task for the nest
176 box requiring birds to initially avoid opaque, and subsequently transparent, flexible barriers to gain
177 access to their nest hole. To do so, birds were required to alter their normal flight path (i.e. level
178 with the nest hole), such that they could enter from the gap under the barrier (Figure 1). The
179 experiment consisted of three phases: 1) habituation to the opaque experimental apparatuses, 2)
180 training phase to go under an opaque barrier, and 3) testing phase with a transparent but otherwise
181 similar barrier. In both years, a Perspex cover (20cmx20cm) was positioned horizontally on top of the
182 box during the training and test phases, to provide birds with experience with transparent plastic (cf.
183 Isaksson et al., 2018; Van Horik et al., 2018) and to act as a cover for the transparent barrier in case
184 of rain (although most experiments occurred when there was no rain).

185 At the start of an experimental session (i.e. when the experimental apparatus was placed on the
186 nest box for each phase), the standard nest box front was replaced with an identical-looking front
187 with integrated infrared and RFID technology that logged visits from birds with or without PIT tags
188 (RFID logger for Schwegler 1B from Biomotors Inc). A panasonic HC camera was mounted on a 50cm
189 high tripod, positioned approximately 10 metres from the nest. Both the nest front and the camera
190 were synchronised and used to quantify the number of times each individual bird landed on/entered
191 the nest hole, and to identify the parents by colour ring and/or plumage characteristics in cases
192 when their PIT tag was not in proximity to the reader (i.e. when they touched the barrier). Each
193 experimental session was presented at the nest box for approximately one hour each day over
194 consecutive days. Videos and infrared/RFID data were reviewed on the same day to determine
195 whether birds progressed to the next experimental phase (see criterion details below). In year one,

196 all approaches during the training phase were scored from video, and in year two, videos during the
197 training phase were used only to confirm RFID data.

198 During the habituation phase, birds were familiarised with the components of the experimental
199 apparatus. These consisted of an opaque rectangular piece of PET plastic film cut from a mobile
200 phone screen protector (0.04W x 10L x 5H cm) covered in camouflage tape attached to the top of
201 the box above the nest hole, a wooden perch positioned at the base of the box attached with metal
202 mesh and wire, and an RFID-equipped nest box front (Figure 1). To pass the habituation phase,
203 individuals had to enter the nest box at least 3 times before moving onto the training phase.

204 During the training phase, the opaque barrier was inverted so that it covered the nest hole (Figure
205 1). This phase ensured birds could perform and were familiar with the motor action of going under
206 the barrier. Birds were visually observed from the camcorder footage, and if they went under the
207 barrier without touching it a minimum of three times, they were advanced to the test phase the
208 following day. During the test phase, the barrier was transparent (i.e. without camouflage tape on
209 the plastic film) (Figure 1). The test phase was performed on consecutive days until birds had made
210 five attempts to enter the nest hole, or until a maximum of six days of experiments (including all
211 phases). These attempts, hereafter referred to as 'trials', were defined as either colliding with or
212 touching the transparent barrier or accessing the nest hole under the barrier without touching it.

213 We measured the wild detour task performance as the proportion of successful trials out of total
214 trials, where higher values indicated better performance. A bird was successful if they flew under the
215 barrier, by flying lower than normal or by landing on the perch/box and jumping under. A bird failed
216 if they touched the barrier either by flying into it, by jumping into it from the perch, or by perching
217 on the side and tapping the edge with their beak. On some occasions the lip of the underside of the
218 barrier touched the bird's back as they jumped under, but this was not considered a fail as it was a
219 clear attempt to avoid the barrier, and likely an artefact of the bird's size rather than a lack of
220 inhibitory control. If a bird was perched at the nest hole and under the barrier, but did not enter the
221 box, they were still scored as having completed a successful trial. Birds had to fly away from the nest
222 box (i.e. out of the camera view) for a new trial to be scored. If a bird repeatedly jumped between
223 the nest hole and the perch (either contacting the barrier or not), their score was based on their first
224 action, which reflects our scoring criteria in the captive task (see below).

225 If both parents did not pass a phase after two days, the experiment was abandoned. If only one
226 parent passed the phase after two days, then the experiment was advanced for the participating
227 parent. Therefore, the experiment took between three to six experimental sessions (i.e. days) to

228 complete in year one (mean =3.7 days \pm 0.13 Standard Error). Experiments started at day ten post-
229 hatching, except for twelve boxes in year one that started at day five-seven due to logistical
230 constraints. In year two the experimental protocol was refined to reduce the number of
231 experimental trials by eliminating a need for a habituation phase, and reducing the criterion for the
232 training phase. Instead of the hour-long habituation phase described above, a dummy apparatus was
233 placed on the box permanently from day eight post-hatching until the start of the training phase.
234 This consisted of wire mesh around the top of the box attached to a plastic rectangle covered with
235 camouflage tape (0.2cm x 6cm x 5cm), and a wooden perch below the nest hole. The criterion for
236 passing the training phase was reduced to one successful approach, instead of three, as our year one
237 data showed that the number of training trials did not influence performance during the test phase
238 (see also results). The experiment took between two and four days to complete in year two (mean =
239 2.4 days \pm 0.12 S.E).

240 In year one, 19 males and 24 females from 29 nests participated in the experiment. The experiment
241 was attempted at three additional nests, but these were excluded because neither parent passed
242 the habituation or training phases. In year two, 21 females and 20 males from 22 nests participated,
243 where at least one parent from all nests reached and participated in the test phase.

244 *Experimental procedures: Detour task in captivity*

245 The captive detour task consisted of the same three experimental phases described above. We
246 piloted different sizes of tubes on a cohort of birds not included in the main analyses and chose a 3.5
247 cm D x 3cm W cylinder tube so that the complexity of the task did not cause ceiling or floor effects in
248 performance caused by the difficulty in detouring around the barrier (Farrar et al., 2020; Völter et
249 al., 2018). A 5 cm high perch was positioned 15 cm in front of the cylinder to standardize the
250 approach direction for each trial (Figure 1). The habituation phase was presented the day after birds
251 arrived in the aviary. Birds had to retrieve a wax worm placed at the edge of an opaque plastic
252 cylinder three times before they received the training phase. Wax worms were euthanised by head
253 compression, so they did not move during the trial. Depending on the bird's progress, the training
254 phase occurred either on the same day, or the following day.

255 In the training phase, birds had to retrieve a wax worm placed in the centre of the same plastic
256 cylinder without touching the exterior of the cylinder four out of five consecutive attempts to
257 retrieve the worm. The test phase was always performed the day after a bird passed the training
258 phase. During this phase birds received ten trials with a transparent cylinder of the same dimensions
259 as the one in the habituation and training phases. To ensure birds did not become sated, half a wax

260 worm was used as a reward during the test trials. A trial was defined as a bird approaching the
261 cylinder and making contact with the barrier (scored as a fail) or retrieving the worm from the side
262 of the tube without making contact with the barrier (scored as a success). The trial ended when the
263 bird retrieved the worm, or flew away from the apparatus, at which point the tube was removed
264 from the testing enclosure, rebaited by the experimenter, and placed back in the enclosure. This
265 procedure was designed such that each approach was measured as a success/fail, as opposed to
266 number of pecks until success, the latter of which may be guided by individual persistence (Van
267 Horik et al., 2018). Allowing birds to consume the worm whether they failed or succeeded at each
268 trial controlled for reward history that may have influenced reinforcement and/or motivation
269 through hunger. All birds ate between eight and ten worms at the test phase, except two birds (four
270 worms, seven worms). We recorded the duration it took each bird to complete ten trials as there
271 was no limit to how long birds had to approach the cylinder for each trial. 35 birds participated in
272 the experiment. One additional bird did not settle in the cage or approach the apparatus and
273 therefore did not participate in the experiment. As for the wild task, we measured the captive
274 detour task performance as the proportion of successful trials out of total trials, such that high
275 values indicated putatively high inhibitory control.

276 *Experimental procedures: Exploration Behaviour*

277 The morning following the birds' arrival to the aviary, we performed an 'exploration in a novel
278 environment' assay, henceforth referred to as exploration behaviour (see also Coomes et al., 2020;
279 adapted from Dingemanse et al., 2002). An access hatch at the back of the bird's cage that led to a
280 larger room (4.60 m W x 3.10 m L x 2.65 m H) was opened. The light in the home cage was turned off
281 and birds were free to enter the room. Once the bird entered the room, the number of hops and
282 flights within and between trees were recorded from the adjacent corridor through one-way glass.
283 'Trees' were made of a wooden upright support and two thick dowels running at right angles to each
284 other (see also Coomes et al., 2020). The trial was complete after 2 minutes, at which point the birds
285 were returned to their home cage. Exploration behaviour was recorded as the sum of the number of
286 hops and flights, and has been shown to be repeatable in our population (O'Shea et al., 2017).

287 *Statistical analysis*

288 All models, unless otherwise specified, were run as Generalised Linear (Mixed) Models in lme4
289 (Bates et al., 2014) in the R statistical software interface (R Core Team, 2014). P values were
290 generated using lmerTest (Kuznetsova et al., 2017). Plots were generated using ggplot (Wilkinson,
291 2011). We used the dredge function from the MuMIn package (Barton 2019) and an information-

292 theoretic approach in combination with model averaging (Grueber et al., 2011). We generated
293 models from a global model from our GLMMs and retained models with an Akaike's Information
294 Criterion corrected for small sample sizes (AICc) within 2 units of the top model. We report the
295 conditional averaged weighted parameter estimates across the retained models. All continuous
296 variables were scaled. We used the vif function in the usdm package (Naimi et al., 2014) to test for
297 collinearity between fixed factors. All variables had Variance Inflation Factor less than 2.5 and
298 therefore were not considered to show multicollinearity. Our R code is included in supplementary
299 materials.

300 *Detour task in the wild*

301 The number of trials undertaken by each bird during the hour test phase varied (mean = 9.6 visits,
302 +/- 0.57 se, 1 min, 26 max). We chose to include a maximum of ten trials as this was consistent with
303 the number of trials in the captive task and existing literature (MacLean et al 2014). We also
304 confirmed that the number of trials used to calculate overall performance did not bias wild detour
305 task performance, if, for example, birds with more trials had higher scores if they learned over
306 successive trials to avoid the barrier. We found that the total number of test trials did not correlate
307 with overall performance using Kendall's tau correlation test from the cor.test function in R ($z=-0.23$,
308 $\tau=-0.02$, $p = 0.79$, $n = 84$). Moreover, for birds that completed at least ten trials, their overall
309 performance calculated from the first five trials was highly correlated with their overall performance
310 calculated from the first ten trials. ($z = 9.42$, $\tau = 0.86$, $p\text{-value} < 0.001$, $n = 69$). Eight individuals
311 completed less than five trials and were included in the analysis (3 birds completed 1 trial, 1 bird
312 completed 2 trials, 1 bird completed 3 trials, and 3 birds completed 4 trials).

313 Initially we examined what fixed effects had a potentially confounding influence on detour task
314 performance in the wild. Lack of any strong effects would lend support for the wild detour task being
315 a reliable test of inhibitory control. It is also important to identify which fixed effects could be driving
316 between individual variation in detour task performance (see repeatability below). We modelled
317 wild detour task performance in a GLMM with a binomial distribution and logit link function, with
318 the number of successes as the numerator and the total trials as a denominator (in R, using cbind,
319 the response variable is entered as two variables, number of successes and number of fails). Our
320 global model included the following fixed effects as potential sources of variation in performance:
321 the number of training trials because the motor action of flying under a barrier could carry-over to
322 the test phase; wing length because size and agility may influence the ability to fly under the barrier;
323 year, lay date and brood size as these may be sources of motivation that may influence parental
324 impulses to feed their chicks; and sex, which has previously been reported to predict inhibitory

325 control performance in a stop-signal task (Lacreuse et al., 2016). Continuous variables were scaled
326 and mean-centred to zero. Site, nest and bird identity (ID) were included as nested random terms.
327 Due to convergence issues associated with overfitting the model with categorical variables, we did
328 not include habitat (conifer versus deciduous) or age in our global model, though visual inspection of
329 these variables and reduced models in which these variables were included suggest they had no
330 effect on performance (Figure S2a, supplementary).

331 *Detour task in captivity*

332 We modelled captive detour task performance in a GLMM with a binomial distribution and logit link
333 function. Our global model included reward history (i.e. number of worms eaten), motivation (i.e.
334 time to complete all ten test trials), habitat, personality, sex and age as fixed factors, and site of
335 capture as a random effect. Habitat was included as a potential ecological confound linked to
336 differences in population density (O'Shea et al., 2018) and/or environmental variability (van Horik et
337 al., 2019). Exploration behaviour was included as a fixed factor as personality may influence how
338 birds engage with the task, or form part of so-called cognitive 'styles' (Sih & Del Giudice, 2012). Sex
339 and age were also included as sources of variation in task performance (Lacreuse et al., 2016;
340 Macdonald et al., 2014). We also investigated whether persistence and captive detour task
341 performance were correlated. To obtain a measure of persistence, we measured the number of
342 times birds pecked at a small transparent case containing a visible but inaccessible mealworm
343 (*Tenebrio molitor*). This data was collected as part of a separate study on foraging choices (Coomes
344 et al., 2020). Persistence data was available for a subset of birds (n=27); therefore, we ran a separate
345 analysis to test for a correlation between the two variables using Kendal's Tau method for non-
346 normal data with the `cor.test` function in R.

347 *Repeatability*

348 We investigated whether individuals showed consistent differences in the wild detour task
349 performance across years (temporal repeatability; but not for the captive task for which we had no
350 repeats). Significant temporal repeatability in performance would suggest that the task measured an
351 intrinsic trait, indicating a permanent environment effect and/or heritability (e.g. Quinn et al., 2009).
352 Our dataset included repeat measures (n=16 observations, 8 individuals), as well as single measures
353 (n= 68) to increase power (Martin et al., 2011). We ran a GLMM as described above, with year as a
354 fixed effect and bird ID as a random effect, and compared this model with another that excluded
355 bird ID using the `anova` function in R. The repeatability estimate was calculated from the variance
356 components and the residual variance as $1/p(1-p)$, where p is the expected probability of success

357 calculated as the mean wild detour task performance in the dataset (Nakagawa et al., 2017). 2.5%
358 and 97.5% Confidence Intervals (CI) were calculated with the function `confint()` using the bootstrap
359 argument with 1000 simulations. Sigma (residual deviation) was estimated to be 1. We also tested
360 whether the inclusion of fixed effects resulted in any change in the repeatability estimate (adjusted
361 repeatability). If the repeatability of wild detour task performance remained significant after
362 inclusion of these effects, this would point to consistent individual differences being explained by
363 inhibitory control.

364 We also estimated contextual repeatability between tasks, significant levels of which would suggest
365 that performance on these tasks could be attributed to a common factor, supporting the hypothesis
366 that these tasks reflect, at least in part, inhibitory control where the prepotent impulse to go directly
367 towards a positive stimulus (either a food reward or a begging offspring), must be inhibited. We ran
368 an additional two GLMMs (with and without bird ID as a random effect) using a dataset that included
369 repeat measures (n = 21 observations, 10 individuals, one of whom was measured both in year 1 and
370 year 2 of the wild task), and singletons (n=98). We included task (wild versus captivity) as a fixed
371 term, and bird ID and site as random terms. We then repeated these analyses to control for fixed
372 effects that were common between tasks and had been retained in the model selection for temporal
373 repeatability, to ensure that any significant repeatability was not driven solely by common factors
374 unrelated to inhibitory control.

375 Research and Animal Ethics. This study was conducted under licences from the Health Products
376 Regulatory Authority (AE19130_P017), The National Parks and Wildlife Services (C11/2017,
377 004/2017, C02/2018 and 001/2018) and the British Trust for Ornithology. The research project
378 received ethical approval from the Animal Welfare Body at University College Cork, and was in
379 accordance with the Guidelines for the Treatment of Animals in Behavioural Research and Teaching
380 (2020).

381 **Results**

382 *Detour task in the wild*

383 The number of times individuals successfully went under the opaque barrier during the training
384 phase varied across individuals (mean = 7.11 ± 0.59 , min 1, max 37) showing that they passed this
385 stage of the task. In year one, during the training phase there were 8 instances from 7 individuals
386 when birds made contact with the opaque barrier, whereas during the test phase there were 130
387 instances from 36 individuals when birds made contact with the transparent barrier (mean =
388 3.02 ± 0.45 SE), confirming that the transparent barrier evoked the prepotent response of flying

389 straight to the nest hole (cf. Van Horik et al., 2018). The mean \pm SE wild detour task performance was
390 0.56 \pm 0.05 SE in year one, and 0.62 \pm 0.04 in year two, suggesting no difference between years. Lay
391 date, sex and wing length were retained in the top models, but evidence that any of these terms had
392 an effect on detour task performance was weak because none of them were statistically significant
393 (Figure 2, Table 1), although males performed marginally worse than females, and there was a
394 tendency for performance to decline with lay date. The number of training trials, brood size and year
395 were not retained in any of the top models, and did not predict task performance (Figure 2, Global
396 model test statistics are provided in Supplementary Table 2).

397 *Detour task in captivity*

398 The mean \pm SE captive detour task performance was 0.41 \pm 0.04 SE, substantially lower than that
399 observed in the wild task. The time it took birds to complete the test phase, the number of worms
400 eaten, sex and exploration behaviour were the retained fixed terms in the top models, but evidence
401 that any of these variables had an effect was weak because none were statistically significant (Figure
402 S2). Age (Figure S2d) and habitat (Figure S3b) were not retained in any of the top models and did not
403 predict task performance (Global model test statistics are provided in Supplementary Table 2). There
404 was no relationship between captive detour task performance and our independent measure of
405 persistence, the number of pecks birds made to an inaccessible worm in a transparent casing ($z=-$
406 1.06; tau = -0.15; $p = 0.29$, $n = 27$).

407 *Repeatability*

408 The temporal repeatability of wild detour task performance across years was low but significant ($R=$
409 0.19, $p < 0.001$, CI = 0.06, 0.22), and remained significant when controlling for sex, lay date and wing
410 length (adjusted $R = 0.15$, $p < 0.001$, CI = 0.001, 0.18) (Figure 3a, c). Contextual repeatability in
411 performance across tasks was low but significant ($R = 0.18$, $p < 0.001$, CI = 0.09, 0.24), and remained
412 significant when controlling for sex (adjusted $R = 0.17$, $p < 0.001$, CI = 0.001, 0.22) (Figure 3b, d).

413 **Discussion**

414 We show that the detour task is repeatable across years and testing environments (wild versus
415 captivity). Furthermore, we controlled for a range of possible confounding variables, and they did
416 not explain the repeatability of performance. To our knowledge, our study is one of two other
417 known studies that have investigated repeatability of inhibitory control in wild birds (Ashton et al.,
418 2018; Shaw, 2017a), only one of which also reported significant repeatability (Ashton et al., 2018).
419 Moreover, our findings contrast with recent reports that the detour task is confounded by cognitive
420 and/or non-cognitive traits unrelated to inhibitory control (Shaw, 2017b; van Horik et al., 2020; Van

421 Horik et al., 2018). Instead, our study lends support for the detour task as a robust measure of
422 inhibitory control in great tits, at least in the context of inhibiting a motor action. By synthesising
423 our findings across both wild and captive versions of the detour task, we discuss the extent to which
424 we can attribute task performance to inhibitory control.

425 Performance on the wild detour task was repeatable across time, and between different tasks.
426 Although repeatability estimates were low, these results point to an underlying, inherent trait that
427 was consistently measured despite differences in year, season, testing location and apparatus-type.
428 Repeatability sets the upper limit for heritability but does not preclude permanent environment
429 effects driving some or even all the intrinsic differences observed. Nevertheless, genetic pedigree
430 studies have shown that motor inhibition is heritable in birds (Langley et al., 2020), and that
431 executive functioning is the most heritable psychological trait in humans (Friedman et al., 2008).
432 While we acknowledge that the relatively small sample size of within-individual measurements may
433 render parameter estimation less reliable (Nakagawa & Schielzeth, 2010), our repeatability
434 estimates were consistent with previous reports in animal cognition. A meta-analysis of repeatability
435 estimates across a range of different cognitive tasks including inhibitory control, problem solving,
436 discrimination and reversal learning, memory, physical and spatial cognition found low to moderate
437 R values for temporal repeatability (0.15 and 0.28) and contextual repeatability (0.20-0.28)
438 (Cauchoix et al., 2018). The two known detour task studies in the meta-analysis had very opposing
439 results: no repeatability in New Zealand Robins (*Petroica australis*) ($R = 0.002$) (Shaw, 2017b) but
440 high repeatability in Australian Magpies (*Cracticus tibicen*) ($R = 0.80$) (Ashton et al., 2018). This
441 suggests that repeatability of the detour task across species and populations may be considerably
442 variable if temporary environmental effects vary across space and time. The particularly high
443 repeatability reported for the Australian Magpies, for example, is likely due to the repeats being
444 taken just two weeks apart (in comparison to the 12 months in this study) which is likely to inflate
445 differences among individuals due to transient factors (Bell et al., 2009; Cole et al., 2011). Disparities
446 in performance and effect sizes across studies are common in detour task studies, and comparative
447 cognition studies generally (Farrar et al., 2020). Our finding that the same individuals performed
448 better on the wild compared to the captive task highlights the difficulty of labelling 'better' or
449 'poorer' inhibitory control abilities when comparing performance within or between species across
450 variations of the detour task. Nevertheless, the significant repeatability, coupled with the lack of
451 evidence that extraneous factors had any effect on repeatability, suggests an underlying trait that is
452 common across tasks in our system, which we attribute to being the inhibition of a prepotent
453 response/habitual behaviour.

454 Motivation (Shaw, 2017b), persistence (Van Horik et al., 2018), and experience (van Horik et al.,
455 2019, 2020) have been proposed to interfere with estimating individual differences in inhibitory
456 control. The use of food rewards may contribute to motivational effects attributed to between-
457 individual differences in hunger state, body condition and/or food preferences (Cooke et al., 2021;
458 Shaw, 2017b). Placing the detour task in front of the nest box access hole allowed us to measure
459 performance independently of food rewards. Moreover, performance was repeatable despite
460 different reward incentives between wild and captive tasks. Although we expected that all
461 individuals would be highly motivated to participate in and complete the inhibitory control tasks as
462 quickly as possible, we nevertheless expected that motivation could vary depending on the
463 reproductive value and the viability of the parent's offspring. However, we found no evidence for
464 this because performance was unaffected by brood size, brood age and lay date. Approach latencies,
465 or time to complete a task, have been interpreted as proxies for motivation to obtain a reward, but
466 we found no effect of these variables on performance in our captive task, nor did a similar study
467 with common pheasants (*Phasianus colchicus*) (van Horik et al., 2020). Body condition may reflect an
468 individual's energetic state, and has been linked with performance in the detour task in wild New
469 Zealand robins (Shaw, 2017b). We did not test for such an effect in the current study as we did not
470 have an accurate measure of body weight at the time of the experiments. Weights were only taken
471 when birds were handled (i.e. at capture), and can fluctuate as much as 5% for great tits in captivity
472 (personal observation), and while parents are provisioning chicks. Notwithstanding the point that
473 individual differences in motivation could theoretically remain due to intrinsic differences in
474 motivation (Morand-Ferron et al., 2016), even when body condition is the same, future studies could
475 integrate a weighing scale into the perch placed in front of the cylinder or the nest box to obtain
476 real-time weight as a proxy for motivation due to energetic state.

477 Persistence, in which animals continue to attempt to obtain a reward, despite persistent errors that
478 do not lead to a reward, has been suggested to contribute to performance in problem solving tasks
479 (Griffin et al., 2015) as well as the detour task (Van Horik et al., 2018). Using data that overlaps with
480 that used here, we found no evidence that persistence in an independent foraging task was related
481 to captive detour task performance (Coomes et al., 2020). The lack of an effect of persistence may
482 be because we scored the initial response (success/fail) for each independent approach to the
483 barrier to quantify inhibitory control, as opposed to measuring performance as the number of pecks
484 before moving around the barrier, the latter quantification being more sensitive to individual
485 differences in persistence, and very similar to how we measured persistence. Indeed, whether
486 behaviour is repeatable or correlates with other traits can be dependent on subtle differences in
487 how behaviour is measured generally (Carter et al., 2012; Davidson et al., 2018).

488 Between-individual differences in animal personality may influence task performance if personality
489 is intrinsically linked to so-called cognitive “styles” (Sih & Del Giudice, 2012). For example,
490 behaviours commonly attributed to the reactive-proactive animal personality axis include slow
491 exploring, environmentally sensitive individuals on one extreme, and fast exploring, routine forming
492 individuals on the other (Réale et al., 2007). These definitions have many parallels with definitions
493 associated with human inhibition and impulsivity, including ‘an impulsive behaviour with no
494 forethought of consequences’ (Moeller et al., 2001). We found no evidence that exploratory
495 behaviour, a common behaviour associated with the fast-slow exploration personality axis, was
496 associated with performance on the detour task in captivity, thus excluding this as an influence on
497 performance in the detour task. This is consistent with reports in black-capped chickadees (*Poecile*
498 *atricapillus*) (Guillette et al., 2017) and domestic dogs (*Canis lupus familiaris*) (Bray et al., 2015), but
499 not common waxbills (*Estrilda astrild*) (Gomes et al., 2020). We note that our measure of personality
500 is purely an index of the fast-slow personality axis (Bell, 2007), and it may be that specific facets of
501 this axis, for example responsiveness, the quality of exploration (i.e. information gathered) and
502 neophobia, need to be measured in isolation to detect links with inhibitory control. Similarly, other
503 kinds of personality axes, or behavioural variation in general, could play a role. Equally, we also note
504 that our measure of inhibitory control likely only captures one facet of self-regulation but there are
505 many others (e.g. delayed gratification; e.g. Mischel et al., 1989) that themselves may be controlled
506 by distinct but related forms of inhibitory control we measure here. Despite the challenges in teasing
507 apart different elements of inhibitory control, and cognition generally, much of the psychology
508 literature suggests that different measures of inhibitory control are a component of a wider latent
509 cognitive variable, such as general inhibition, executive functioning and general intelligence
510 (Anderson & Weaver, 2009; Aron et al., 2004; Ashton et al., 2018; Bari & Robbins, 2013; Shaw et al.,
511 2015), but it remains to be seen whether that is also true in wild study systems.

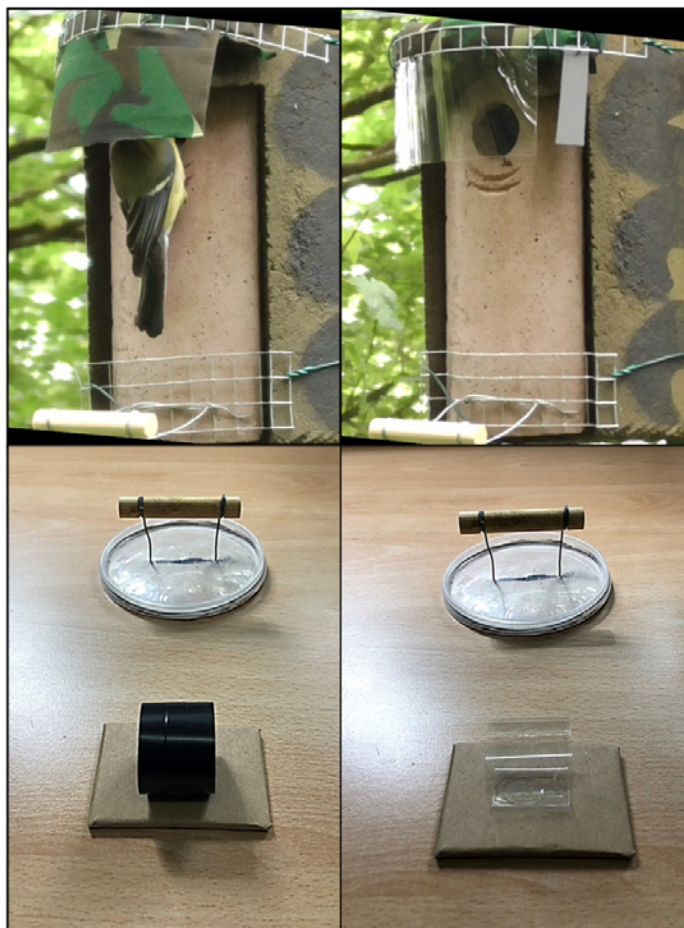
512 In both detour tasks, performance on successive trials could have been driven by motor learning
513 from previous trials. The acquisition of a motor routine has been shown to affect detour task
514 performance in pheasants (van Horik et al., 2020), but this was not the case in our wild task. It is
515 often assumed that a major drawback of cognitive tests in the wild is that it is difficult to control for
516 previous experience, but our results suggest that the wild detour task is not sensitive to differences
517 in the number of training or test trials in which individuals participate. Fine scale local factors could
518 also drive consistency across years - for example approach direction and perception of the
519 transparent barrier associated with nest box orientation or lighting conditions - but many of these
520 are unlikely because birds typically change their nest box every year, and only one of the nest boxes
521 was occupied in both years in our sample. Nevertheless it is highly likely that a range of other

522 confounding effects could have influenced performance in our tasks, some of which are likely to be
523 normally overlooked even in controlled experiments (e.g. Dunn et al., 2011).

524 The field of animal cognition has made major advances in describing cognitive variation between
525 and within species, yet obtaining unbiased and realistic estimates of cognitive variation in natural
526 populations remains a significant challenge (Morand-Ferron et al., 2016; Rowe & Healy, 2014;
527 Thornton et al., 2014). Limited participation in self-administered cognitive trials in the wild
528 potentially leads to bias towards some kinds of individuals, for example those with higher cognitive
529 abilities (Reichert et al., 2020). By deploying a modified version of the detour task that we developed
530 specifically for our system, where birds were compelled to visit, we were able to minimise
531 participation bias at the population level, which can have a big impact on parameter estimation
532 generally. Additionally, our approach limited the effects of human interference and social
533 interactions by conducting the task at isolated locations. Finally, our results support the traditionally
534 held view that the detour task is a reliable measure of inhibitory control, a cognitive process that is
535 likely an important driver of functionally important behavioural plasticity. While it may never be
536 feasible to study inhibitory control as a discrete module in isolation from extraneous/integrated
537 processes, which is true for many cognitive processes, it may not be advisable or necessary to do so
538 when addressing questions in evolutionary ecology (Morand-Ferron & Quinn, 2015) since selection
539 rarely acts on individual genes. Overall, cognitive estimates derived from the detour task can hold
540 value either as a stand-alone task specifically measuring inhibition, or as part of a larger test battery
541 aimed at understanding general cognitive ability.

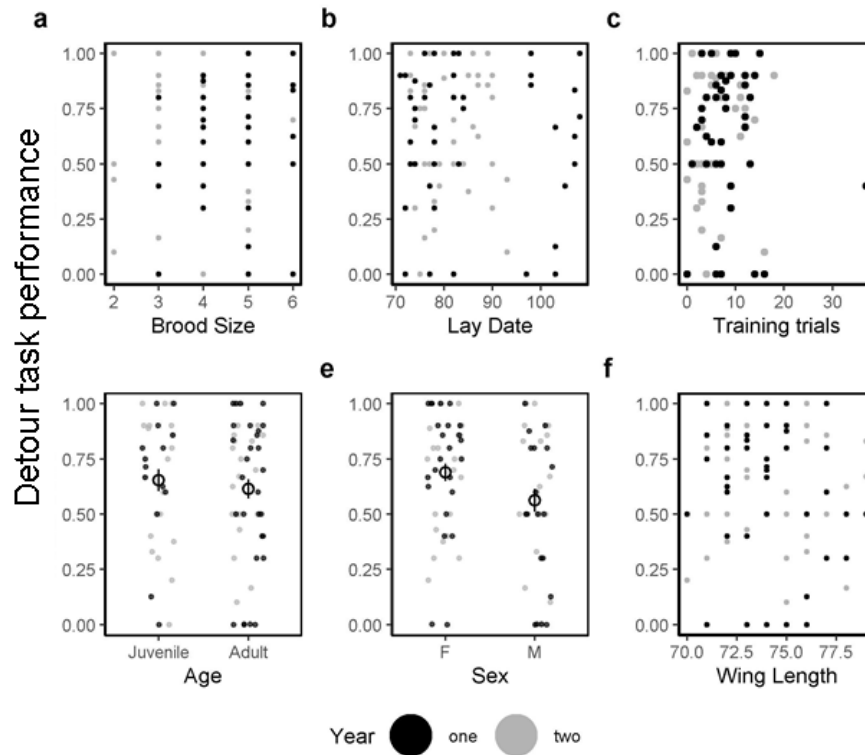
542 Author contribution: G.L.D. designed the experiment with input from all authors. G.L.D. ran the
543 experiments with assistance for M.S.R., J.R.C., I.G.K and I.D.H. G.L.D analysed the data. G.L.D and
544 J.L.Q wrote the manuscript with input from all authors.

545 Acknowledgements: We wish to thank Sam Bayley, Jodie Crane, and James O'Neill for assistance
546 with fieldwork, and Amy Cooke for assistance in the aviary. Thank you to Karen Cogan, Allan
547 Whitaker and Luke Harman for technical support. Funding came from the European Research
548 Council under the European Union's Horizon 2020 Programme (FP7/2007-2013)/ERC Consolidator
549 Grant "EVOLCOCOG" Project No. 617509, awarded to J.L.Q., and from a Science Foundation Ireland
550 ERC Support Grant 14/ERC/B3118 to J.L.Q.



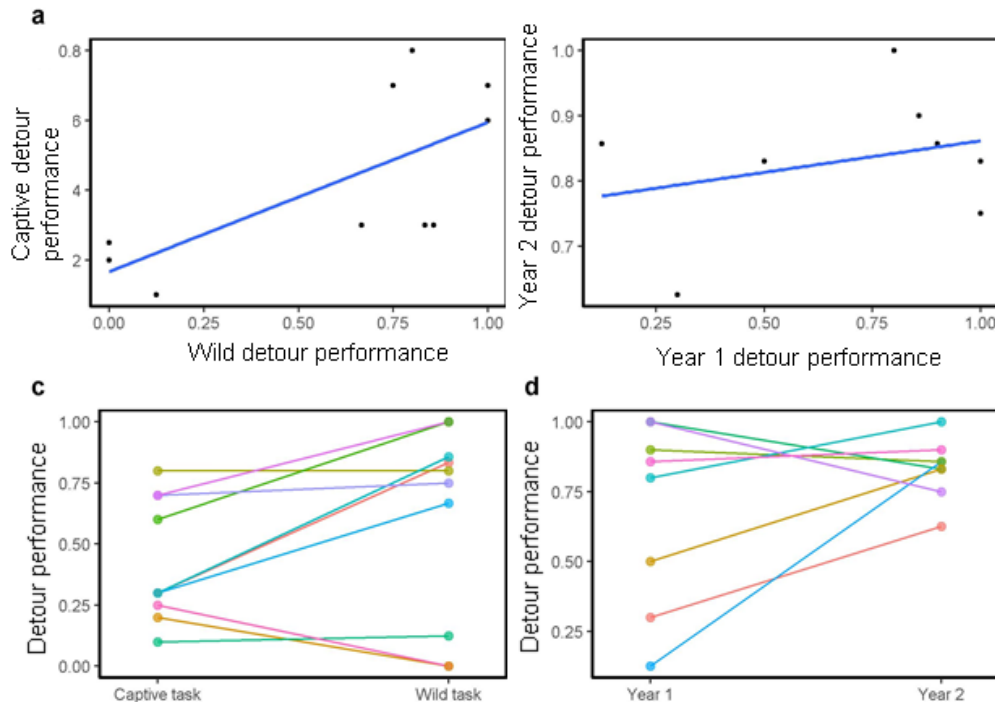
551

552 Figure 1. The phases of the detour task in the wild (top) and in captivity (bottom). Training phases
553 with an opaque barrier (left) and test phases with a transparent barrier (right). Perches were
554 positioned below the nest box entrance hole, and in front of the cylinders. For the habituation phase
555 in the wild, the opaque barrier was inverted 180° so it was positioned above the box. For the
556 habituation phase in captivity, the reward was placed visibly on the side of the tube.



557

558 Figure 2. Detour performance (the proportion of successful trials out of total trials, where higher
559 values indicate better performance) in the wild for (a) brood size, (b) lay date (days since 1st March),
560 (c) total number of training trials, (d) age, (e) sex (F=female, M=male), (f) wing length (mm). Points
561 represent individuals where black points were taken in year one, and grey in year two. Datapoints
562 for c, d and f are jittered along the horizontal axis to reduce overlap. Open circle and line represent
563 mean ± standard error across both years. Lay date and sex tended towards significance, and wing
564 length was retained in the final model (Table1).



565

566 Figure 3. The relationship between individual detour task performance (the proportion of successful
 567 trials out of total trials, where higher values indicate better performance) between (a) the wild and
 568 captive detour tasks, and (b) year 1 and year 2 wild detour tasks. Reaction norms for (c) contextual
 569 repeatability across captive and wild detour tasks, and (d) temporal repeatability across years with
 570 the wild detour task. Circles represent individual performance, a regression line is fitted in (a) and
 571 (b), and lines (c) and (d) connect repeated measures within individuals in (c) and (d). Lines and circles
 572 are in different colours per individual. Repeatability was significant for both analyses, indicating that
 573 rank order differences across tasks are more similar than expected by chance.

574 Table 1. Binomial GLMMs of detour task performance (proportion of successful test trials) for the
 575 wild detour task (n=84) and the captive detour task (n=35). The values shown are the conditional
 576 average of the top models within two AICc of the best model. All continuous variables were scaled.
 577 Reference categories for binary variables are in brackets.

	Estimate ± Standard Error	z	p value
Wild detour task			
Intercept	0.96±0.38	2.54	0.01
Lay date	-0.34±0.20	-1.65	0.10
Sex (Female)	-0.867±0.49	-1.73	0.08
Wing length	0.35±0.27	1.27	0.20
Captive detour task			

Intercept	-0.32±0.20	-1.57	0.11
Time to complete test	0.14±0.12	1.23	0.25
Sex (Female)	-0.24±0.24	-0.96	0.34
Exploration behaviour	0.09±0.13	0.66	0.51
Total rewards eaten	0.19±0.13	1.36	0.17

578

579 REFERENCES

580 Amici, F., Aureli, F., & Call, J. (2013). Fission-Fusion Dynamics, Behavioral Flexibility, and Inhibitory
581 Control in Primates. *Current Biology*, 23(13). <https://doi.org/10.1016/j.cub.2013.06.020>

582 Amici, F., Call, J., Watzek, J., Brosnan, S., & Aureli, F. (2018). Social inhibition and behavioural
583 flexibility when the context changes: A comparison across six primate species. *Scientific*
584 *Reports*, 8(1). <https://doi.org/10.1038/s41598-018-21496-6>

585 Anderson, M. C., & Weaver, C. (2009). Inhibitory Control over Action and Memory. In *Encyclopedia of*
586 *Neuroscience*. <https://doi.org/10.1016/B978-008045046-9.00421-6>

587 Aplin, L. M., Farine, D. R., Morand-Ferron, J., Cockburn, A., Thornton, A., & Sheldon, B. C. (2015).
588 Experimentally induced innovations lead to persistent culture via conformity in wild birds.
589 *Nature*, 518(7540). <https://doi.org/10.1038/nature13998>

590 Aron, A. R., Robbins, T. W., & Poldrack, R. A. (2004). Inhibition and the right inferior frontal cortex. In
591 *Trends in Cognitive Sciences* (Vol. 8, Issue 4). <https://doi.org/10.1016/j.tics.2004.02.010>

592 Ashton, B. J., Ridley, A. R., Edwards, E. K., & Thornton, A. (2018). Cognitive performance is linked to
593 group size and affects fitness in Australian magpies. *Nature*, 554(7692).
594 <https://doi.org/10.1038/nature25503>

595 Bari, A., & Robbins, T. W. (2013). Inhibition and impulsivity: Behavioral and neural basis of response
596 control. In *Progress in Neurobiology* (Vol. 108).
597 <https://doi.org/10.1016/j.pneurobio.2013.06.005>

598 Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using
599 Eigen and S4. R package version 1.1-7, <http://CRAN.R-project.org/package=lme4>. *R Package*
600 *Version*.

601 Bell, A. M. (2007). Future directions in behavioural syndromes research. In *Proceedings of the Royal*
602 *Society B: Biological Sciences* (Vol. 274, Issue 1611). <https://doi.org/10.1098/rspb.2006.0199>

- 603 Bell, A. M., Hankison, S. J., & Laskowski, K. L. (2009). The repeatability of behaviour: a meta-analysis.
604 In *Animal Behaviour* (Vol. 77, Issue 4). <https://doi.org/10.1016/j.anbehav.2008.12.022>
- 605 Benson-Amram, S., Weldele, M. L., & Holekamp, K. E. (2013). A comparison of innovative problem-
606 solving abilities between wild and captive spotted hyaenas, *Crocuta crocuta*. *Animal Behaviour*,
607 85(2). <https://doi.org/10.1016/j.anbehav.2012.11.003>
- 608 Bray, E. E., MacLean, E. L., & Hare, B. A. (2015). Increasing arousal enhances inhibitory control in
609 calm but not excitable dogs. *Animal Cognition*, 18(6). [https://doi.org/10.1007/s10071-015-](https://doi.org/10.1007/s10071-015-0901-1)
610 0901-1
- 611 Butler, S. J., Whittingham, M. J., Quinn, J. L., & Cresswell, W. (2006). Time in captivity, individual
612 differences and foraging behaviour in wild-caught chaffinches. *Behaviour*, 143(4).
613 <https://doi.org/10.1163/156853906776240632>
- 614 Carter, A. J., Marshall, H. H., Heinsohn, R., & Cowlshaw, G. (2012). How not to measure boldness:
615 Novel object and antipredator responses are not the same in wild baboons. *Animal Behaviour*,
616 84(3). <https://doi.org/10.1016/j.anbehav.2012.06.015>
- 617 Cauchard, L., Boogert, N. J., Lefebvre, L., Dubois, F., & Doligez, B. (2013). Problem-solving
618 performance is correlated with reproductive success in a wild bird population. *Animal*
619 *Behaviour*, 85(1). <https://doi.org/10.1016/j.anbehav.2012.10.005>
- 620 Cauchoix, M., Chow, P. K. Y., Van Horik, J. O., Atance, C. M., Barbeau, E. J., Barragan-Jason, G., Bize,
621 P., Boussard, A., Buechel, S. D., Cabirol, A., Cauchard, L., Claidière, N., Dalesman, S., Devaud, J.,
622 M., Didic, M., Doligez, B., Fagot, J., Fichtel, C., Henke-Von Der Malsburg, J., ... Morand-Ferron, J.
623 (2018). The repeatability of cognitive performance: A meta-analysis. *Philosophical Transactions*
624 *of the Royal Society B: Biological Sciences*, 373(1756). <https://doi.org/10.1098/rstb.2017.0281>
- 625 Cauchoix, M., Hermer, E., Chaîne, A. S., & Morand-Ferron, J. (2017). Cognition in the field:
626 Comparison of reversal learning performance in captive and wild passerines. *Scientific Reports*,
627 7(1). <https://doi.org/10.1038/s41598-017-13179-5>
- 628 Cole, E. F., Cram, D. L., & Quinn, J. L. (2011). Individual variation in spontaneous problem-solving
629 performance among wild great tits. *Animal Behaviour*, 81(2).
630 <https://doi.org/10.1016/j.anbehav.2010.11.025>
- 631 Cole, E. F., Morand-Ferron, J., Hinks, A. E., & Quinn, J. L. (2012). Cognitive ability influences
632 reproductive life history variation in the wild. *Current Biology*, 22(19).

- 633 <https://doi.org/10.1016/j.cub.2012.07.051>
- 634 Cole, E. F., & Quinn, J. L. (2012). Personality and problem-solving performance explain competitive
635 ability in the wild. *Proceedings of the Royal Society B: Biological Sciences*, 279(1731).
636 <https://doi.org/10.1098/rspb.2011.1539>
- 637 Cooke, A. C., Davidson, G. L., van Oers, K., & Quinn, J. L. (2021). Motivation, accuracy and positive
638 feedback through experience explain innovative problem solving and its repeatability. *Animal*
639 *Behaviour*, 174. <https://doi.org/10.1016/j.anbehav.2021.01.024>
- 640 Coomes, J. R., Davidson, G. L., Reichert, M. S., Kulahci, I. G., Troisi, C. A., & Quinn, J. L. (2020).
641 Inhibitory control, personality, and manipulated ecological conditions influence foraging
642 plasticity in the great tit. In *bioRxiv*. <https://doi.org/10.1101/2020.12.16.423008>
- 643 Davidson, G. L., Reichert, M. S., Crane, J. M. S., O'Shea, W., & Quinn, J. L. (2018). Repeatable aversion
644 across threat types is linked with life-history traits but is dependent on how aversion is
645 measured. *Royal Society Open Science*, 5(2). <https://doi.org/10.1098/rsos.172218>
- 646 Diamond, A. (1990). Developmental Time Course in Human Infants and Infant Monkeys, and the
647 Neural Bases of, Inhibitory Control in Reaching. *Annals of the New York Academy of Sciences*,
648 608(1), 637–676. <https://doi.org/10.1111/j.1749-6632.1990.tb48913.x>
- 649 Dingemanse, N. J., Both, C., Drent, P. J., Van Oers, K., & Van Noordwijk, A. J. (2002). Repeatability
650 and heritability of exploratory behaviour in great tits from the wild. *Animal Behaviour*, 64(6).
651 <https://doi.org/10.1006/anbe.2002.2006>
- 652 Dohm, M. R. (2002). Repeatability estimates do not always set an upper limit to heritability.
653 *Functional Ecology*, 16(2). <https://doi.org/10.1046/j.1365-2435.2002.00621.x>
- 654 Dougherty, L. R., & Guillette, L. M. (2018). Linking personality and cognition: A meta-analysis. In
655 *Philosophical Transactions of the Royal Society B: Biological Sciences* (Vol. 373, Issue 1756).
656 <https://doi.org/10.1098/rstb.2017.0282>
- 657 Dunn, J. C., Cole, E. F., & Quinn, J. L. (2011). Personality and parasites: Sex-dependent associations
658 between avian malaria infection and multiple behavioural traits. *Behavioral Ecology and*
659 *Sociobiology*, 65(7). <https://doi.org/10.1007/s00265-011-1156-8>
- 660 Epstein, L. H., Dearing, K. K., Temple, J. L., & Cavanaugh, M. D. (2008). Food reinforcement and
661 impulsivity in overweight children and their parents. *Eating Behaviors*, 9(3).
662 <https://doi.org/10.1016/j.eatbeh.2007.10.007>

- 663 Farrar, B. G., Voudouris, K., & Clayton, N. S. (2020). Replications, comparisons, sampling and the
664 problem of representativeness in animal behavior and cognition research. *Psyarxiv, August*.
665 <https://psyarxiv.com/2vt4k/>
- 666 Forss, S. I. F., Schuppli, C., Haiden, D., Zweifel, N., & van Schaik, C. P. (2015). Contrasting responses to
667 novelty by wild and captive orangutans. *American Journal of Primatology, 77*(10).
668 <https://doi.org/10.1002/ajp.22445>
- 669 Friedman, N. P., Miyake, A., Young, S. E., DeFries, J. C., Corley, R. P., & Hewitt, J. K. (2008). Individual
670 Differences in Executive Functions Are Almost Entirely Genetic in Origin. *Journal of*
671 *Experimental Psychology: General, 137*(2). <https://doi.org/10.1037/0096-3445.137.2.201>
- 672 Gomes, A. C. R., Guerra, S., Silva, P. A., Marques, C. I., Trigo, S., Boogert, N. J., & Cardoso, G. C.
673 (2020). Proactive common waxbills make fewer mistakes in a cognitive assay, the detour-
674 reaching task. *Behavioral Ecology and Sociobiology, 74*(3). [https://doi.org/10.1007/s00265-](https://doi.org/10.1007/s00265-020-2809-2)
675 [020-2809-2](https://doi.org/10.1007/s00265-020-2809-2)
- 676 Griffin, A. S., Guillette, L. M., & Healy, S. D. (2015). Cognition and personality: An analysis of an
677 emerging field. In *Trends in Ecology and Evolution* (Vol. 30, Issue 4).
678 <https://doi.org/10.1016/j.tree.2015.01.012>
- 679 Grueber, C. E., Nakagawa, S., Laws, R. J., & Jamieson, I. G. (2011). Multimodel inference in ecology
680 and evolution: Challenges and solutions. In *Journal of Evolutionary Biology* (Vol. 24, Issue 4).
681 <https://doi.org/10.1111/j.1420-9101.2010.02210.x>
- 682 Guidelines for the treatment of animals in behavioural research and teaching. (2020). *Animal*
683 *Behaviour, 159*. <https://doi.org/10.1016/j.anbehav.2019.11.002>
- 684 Guillette, L. M., Naguib, M., & Griffin, A. S. (2017). Individual differences in cognition and personality.
685 In *Behavioural Processes* (Vol. 134). <https://doi.org/10.1016/j.beproc.2016.12.001>
- 686 Hampshire, A., & Sharp, D. J. (2015). Contrasting network and modular perspectives on inhibitory
687 control. In *Trends in Cognitive Sciences* (Vol. 19, Issue 8).
688 <https://doi.org/10.1016/j.tics.2015.06.006>
- 689 Houben, K., Nederkoorn, C., & Jansen, A. (2014). Eating on impulse: The relation between
690 overweight and food-specific inhibitory control. *Obesity, 22*(5).
691 <https://doi.org/10.1002/oby.20670>
- 692 Isaksson, E., Utku Urhan, A., & Brodin, A. (2018). High level of self-control ability in a small passerine

- 693 bird. *Behavioral Ecology and Sociobiology*, 72(7). <https://doi.org/10.1007/s00265-018-2529-z>
- 694 Jelbert, S. A., Taylor, A. H., & Gray, R. D. (2016). Does absolute brain size really predict self-control?
695 Hand-tracking training improves performance on the A-not-B task. *Biology Letters*, 12(2).
696 <https://doi.org/10.1098/rsbl.2015.0871>
- 697 Jentsch, J. D., & Taylor, J. R. (1999). Impulsivity resulting from frontostriatal dysfunction in drug
698 abuse: Implications for the control of behavior by reward-related stimuli. In
699 *Psychopharmacology* (Vol. 146, Issue 4). <https://doi.org/10.1007/PL00005483>
- 700 Johnson-Ulrich, L., Holekamp, K. E., & Hambrick, D. Z. (2020). Innovative problem-solving in wild
701 hyenas is reliable across time and contexts. *Scientific Reports*, 10(1).
702 <https://doi.org/10.1038/s41598-020-69953-5>
- 703 Kabadayi, C., Bobrowicz, K., & Osvath, M. (2018). The detour paradigm in animal cognition. In *Animal*
704 *Cognition* (Vol. 21, Issue 1). <https://doi.org/10.1007/s10071-017-1152-0>
- 705 Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear
706 Mixed Effects Models. *Journal of Statistical Software*, 82(13).
707 <https://doi.org/10.18637/jss.v082.i13>
- 708 Lacreuse, A., Gullstrand, J., & Fagot, J. (2016). Sex differences in inhibitory control in socially-housed
709 baboons (*Papio papio*). *Behavioural Brain Research*, 312, 231–237.
710 <https://doi.org/10.1016/j.bbr.2016.06.032>
- 711 Langley, E. J. G., Adams, G., Beardsworth, C. E., Dawson, D. A., Laker, P. R., Van Horik, J. O.,
712 Whiteside, M. A., Wilson, A. J., & Madden, J. R. (2020). Heritability and correlations among
713 learning and inhibitory control traits. *Behavioral Ecology*, 31(3).
714 <https://doi.org/10.1093/beheco/araa029>
- 715 Loeber, S., & Duka, T. (2009). Acute alcohol impairs conditioning of a behavioural reward-seeking
716 response and inhibitory control processes - Implications for addictive disorders. *Addiction*,
717 104(12). <https://doi.org/10.1111/j.1360-0443.2009.02718.x>
- 718 Macdonald, J. A., Beauchamp, M. H., Crigan, J. A., & Anderson, P. J. (2014). Age-related differences in
719 inhibitory control in the early school years. *Child Neuropsychology*, 20(5).
720 <https://doi.org/10.1080/09297049.2013.822060>
- 721 MacLean, E. L., Hare, B., Nun, C. L., Adress, E., Amic, F., Anderson, R. C., Aureli, F., Baker, J. M.,
722 Bania, A. E., Barnard, A. M., Boogert, N. J., Brannon, E. M., Bray, E. E., Bray, J., Brent, L. J. N.,

- 723 Burkart, J. M., Call, J., Cantlo, J. F., Chek, L. G., ... Zhao, Y. (2014). The evolution of self-control.
724 *Proceedings of the National Academy of Sciences of the United States of America*, *111*(20).
725 <https://doi.org/10.1073/pnas.1323533111>
- 726 Martin, J. G. A., Nussey, D. H., Wilson, A. J., & Réale, D. (2011). Measuring individual differences in
727 reaction norms in field and experimental studies: A power analysis of random regression
728 models. *Methods in Ecology and Evolution*, *2*(4). <https://doi.org/10.1111/j.2041->
729 [210X.2010.00084.x](https://doi.org/10.1111/j.2041-210X.2010.00084.x)
- 730 McCune, K. B., Jablonski, P., Lee, S. I., & Ha, R. R. (2019). Captive jays exhibit reduced problem-
731 solving performance compared to wild conspecifics. *Royal Society Open Science*, *6*(1).
732 <https://doi.org/10.1098/rsos.181311>
- 733 Mischel, W., Shoda, Y., & Rodriguez, M. L. (1989). Delay of gratification in children. *Science*,
734 *244*(4907). <https://doi.org/10.1126/science.2658056>
- 735 Moeller, F. G., Barratt, E. S., Dougherty, D. M., Schmitz, J. M., & Swann, A. C. (2001). Psychiatric
736 aspects of impulsivity. In *American Journal of Psychiatry* (Vol. 158, Issue 11).
737 <https://doi.org/10.1176/appi.ajp.158.11.1783>
- 738 Morand-Ferron, J., Cole, E. F., & Quinn, J. L. (2016). Studying the evolutionary ecology of cognition in
739 the wild: A review of practical and conceptual challenges. *Biological Reviews*, *91*(2).
740 <https://doi.org/10.1111/brv.12174>
- 741 Morand-Ferron, J., Cole, E. F., Rawles, J. E. C., & Quinn, J. L. (2011). Who are the innovators? A field
742 experiment with 2 passerine species. *Behavioral Ecology*, *22*(6).
743 <https://doi.org/10.1093/beheco/arr120>
- 744 Morand-Ferron, J., Hamblin, S., Cole, E. F., Aplin, L. M., & Quinn, J. L. (2015). Taking the operant
745 paradigm into the field: Associative learning in wild great tits. *PLoS ONE*, *10*(8).
746 <https://doi.org/10.1371/journal.pone.0133821>
- 747 Morand-Ferron, J., & Quinn, J. L. (2011). Larger groups of passerines are more efficient problem
748 solvers in the wild. *Proceedings of the National Academy of Sciences of the United States of*
749 *America*, *108*(38). <https://doi.org/10.1073/pnas.1111560108>
- 750 Morand-Ferron, J., & Quinn, J. L. (2015). The evolution of cognition in natural populations. In *Trends*
751 *in Cognitive Sciences* (Vol. 19, Issue 5). <https://doi.org/10.1016/j.tics.2015.03.005>
- 752 Mouchet, A., & Dingemanse, N. J. (2021). A quantitative genetics approach to validate lab-versus

- 753 field-based behavior in novel environments. *Behavioral Ecology*.
754 <https://doi.org/10.1093/beheco/arab059>
- 755 Muth, F., Cooper, T. R., Bonilla, R. F., & Leonard, A. S. (2018). A novel protocol for studying bee
756 cognition in the wild. *Methods in Ecology and Evolution*, 9(1). [https://doi.org/10.1111/2041-](https://doi.org/10.1111/2041-210X.12852)
757 210X.12852
- 758 Naimi, B., Hamm, N. A. S., Groen, T. A., Skidmore, A. K., & Toxopeus, A. G. (2014). Where is positional
759 uncertainty a problem for species distribution modelling? *Ecography*, 37(2).
760 <https://doi.org/10.1111/j.1600-0587.2013.00205.x>
- 761 Nakagawa, S., Johnson, P. C. D., & Schielzeth, H. (2017). The coefficient of determination R² and
762 intra-class correlation coefficient from generalized linear mixed-effects models revisited and
763 expanded. *Journal of the Royal Society Interface*, 14(134).
764 <https://doi.org/10.1098/rsif.2017.0213>
- 765 Nakagawa, S., & Schielzeth, H. (2010). Repeatability for Gaussian and non-Gaussian data: A practical
766 guide for biologists. In *Biological Reviews* (Vol. 85, Issue 4). [https://doi.org/10.1111/j.1469-](https://doi.org/10.1111/j.1469-185X.2010.00141.x)
767 185X.2010.00141.x
- 768 O'Shea, W., O'Halloran, J., & Quinn, J. L. (2018). Breeding phenology, provisioning behaviour, and
769 unusual patterns of life history variation across an anthropogenic heterogeneous landscape.
770 *Oecologia*, 188(4). <https://doi.org/10.1007/s00442-018-4155-x>
- 771 O'Shea, W., Serrano-Davies, E., & Quinn, J. L. (2017). Do personality and innovativeness influence
772 competitive ability? An experimental test in the great tit. *Behavioral Ecology*, 28(6).
773 <https://doi.org/10.1093/beheco/arx104>
- 774 Quinn, J. L., Patrick, S. C., Bouwhuis, S., Wilkin, T. A., & Sheldon, B. C. (2009). Heterogeneous
775 selection on a heritable temperament trait in a variable environment. *Journal of Animal*
776 *Ecology*, 78(6). <https://doi.org/10.1111/j.1365-2656.2009.01585.x>
- 777 R Core Team. (2014). R Core Team (2014). R: A language and environment for statistical computing.
778 *R Foundation for Statistical Computing, Vienna, Austria. URL Http://Www.R-Project.Org/*.
- 779 Réale, D., Reader, S. M., Sol, D., McDougall, P. T., & Dingemanse, N. J. (2007). Integrating animal
780 temperament within ecology and evolution. In *Biological Reviews* (Vol. 82, Issue 2).
781 <https://doi.org/10.1111/j.1469-185X.2007.00010.x>
- 782 Reddy, R. B., MacLean, E. L., Sandel, A. A., & Hare, B. (2015). Social inhibitory control in five lemur

- 783 species. *Primates*, 56(3). <https://doi.org/10.1007/s10329-015-0467-1>
- 784 Reichert, M. S., Crofts, S. J., Davidson, G. L., Firth, J. A., Kulahci, I. G., & Quinn, J. L. (2020). Multiple
785 factors affect discrimination learning performance, but not between-individual variation, in
786 wild mixed-species flocks of birds. *Royal Society Open Science*, 7(4).
787 <https://doi.org/10.1098/rsos.192107>
- 788 Rowe, C., & Healy, S. D. (2014). Measuring variation in cognition. In *Behavioral Ecology* (Vol. 25, Issue
789 6). <https://doi.org/10.1093/beheco/aru090>
- 790 Shaw, R. C. (2017a). Testing cognition in the wild: factors affecting performance and individual
791 consistency in two measures of avian cognition. In *Behavioural Processes* (Vol. 134).
792 <https://doi.org/10.1016/j.beproc.2016.06.004>
- 793 Shaw, R. C. (2017b). Testing cognition in the wild: factors affecting performance and individual
794 consistency in two measures of avian cognition. In *Behavioural Processes* (Vol. 134, pp. 31–36).
795 Elsevier B.V. <https://doi.org/10.1016/j.beproc.2016.06.004>
- 796 Shaw, R. C., Boogert, N. J., Clayton, N. S., & Burns, K. C. (2015). Wild psychometrics: Evidence for
797 “general” cognitive performance in wild New Zealand robins, *Petroica longipes*. *Animal*
798 *Behaviour*, 109. <https://doi.org/10.1016/j.anbehav.2015.08.001>
- 799 Shaw, R. C., & Schmelz, M. (2017). Cognitive test batteries in animal cognition research: evaluating
800 the past, present and future of comparative psychometrics. In *Animal Cognition* (Vol. 20, Issue
801 6). <https://doi.org/10.1007/s10071-017-1135-1>
- 802 Sih, A., & Del Giudice, M. (2012). Linking behavioural syndromes and cognition: A behavioural
803 ecology perspective. In *Philosophical Transactions of the Royal Society B: Biological Sciences*
804 (Vol. 367, Issue 1603). <https://doi.org/10.1098/rstb.2012.0216>
- 805 Sonnenberg, B. R., Branch, C. L., Pitera, A. M., Bridge, E., & Pravosudov, V. V. (2019). Natural
806 Selection and Spatial Cognition in Wild Food-Caching Mountain Chickadees. *Current Biology*,
807 29(4). <https://doi.org/10.1016/j.cub.2019.01.006>
- 808 Tang, D., & Schmeichel, B. J. (2014). Stopping anger and anxiety: Evidence that inhibitory ability
809 predicts negative emotional responding. *Cognition and Emotion*, 28(1).
810 <https://doi.org/10.1080/02699931.2013.799459>
- 811 Thornton, A., Isden, J., & Madden, J. R. (2014). Toward wild psychometrics: Linking individual
812 cognitive differences to fitness. *Behavioral Ecology*, 25(6).

813 <https://doi.org/10.1093/beheco/aru095>

814 Toledo, S., Shohami, D., Schiffner, I., Lourie, E., Orchan, Y., Bartan, Y., & Nathan, R. (2020). Cognitive
815 map-based navigation in wild bats revealed by a new high-throughput tracking system. *Science*,
816 *369*(6500). <https://doi.org/10.1126/science.aax6904>

817 Troisi, C. A., Cooke, A. C., Davidson, G. L., de la Hera, I., Reichert, M. S., & Quinn, J. L. (2020). No
818 evidence for cross-contextual consistency in spatial learning and behavioural flexibility in a
819 passerine. In *bioRxiv*. <https://doi.org/10.1101/2020.09.04.282566>

820 van Horik, J. O., Beardsworth, C. E., Laker, P. R., Langley, E. J. G., Whiteside, M. A., & Madden, J. R.
821 (2019). Unpredictable environments enhance inhibitory control in pheasants. *Animal Cognition*,
822 *22*(6). <https://doi.org/10.1007/s10071-019-01302-0>

823 van Horik, J. O., Beardsworth, C. E., Laker, P. R., Whiteside, M. A., & Madden, J. R. (2020). Response
824 learning confounds assays of inhibitory control on detour tasks. *Animal Cognition*, *23*(1).
825 <https://doi.org/10.1007/s10071-019-01330-w>

826 Van Horik, J. O., Langley, E. J. G., Whiteside, M. A., Laker, P. R., Beardsworth, C. E., & Madden, J. R.
827 (2018). Do detour tasks provide accurate assays of inhibitory control? *Proceedings of the Royal*
828 *Society B: Biological Sciences*, *285*(1875). <https://doi.org/10.1098/rspb.2018.0150>

829 Völter, C. J., Tinklenberg, B., Call, J., & Seed, A. M. (2018). Comparative psychometrics: Establishing
830 what differs is central to understanding what evolves. In *Philosophical Transactions of the Royal*
831 *Society B: Biological Sciences* (Vol. 373, Issue 1756). <https://doi.org/10.1098/rstb.2017.0283>

832 Wilkinson, L. (2011). ggplot2: Elegant Graphics for Data Analysis by WICKHAM, H. *Biometrics*, *67*(2).
833 <https://doi.org/10.1111/j.1541-0420.2011.01616.x>

834 Wilson, A. J. (2018). How should we interpret estimates of individual repeatability? *Evolution Letters*,
835 *2*(1). <https://doi.org/10.1002/evl3.40>

836

837