



# Royal Netherlands Academy of Arts and Sciences (KNAW) KONINKLIJKE NEDERLANDSE AKADEMIE VAN WETENSCHAPPEN

## Motivation, accuracy and positive feedback through experience explain innovative problem solving and its repeatability

Cooke, Amy C.; Davidson, Gabrielle L.; van Oers, Kees; Quinn, John L.

### **published in**

Animal Behaviour  
2021

### **DOI (link to publisher)**

[10.1016/j.anbehav.2021.01.024](https://doi.org/10.1016/j.anbehav.2021.01.024)

### **document version**

Peer reviewed version

### **document license**

CC BY-NC-ND

[Link to publication in KNAW Research Portal](#)

### **citation for published version (APA)**

Cooke, A. C., Davidson, G. L., van Oers, K., & Quinn, J. L. (2021). Motivation, accuracy and positive feedback through experience explain innovative problem solving and its repeatability. *Animal Behaviour*, 174, 249-261. <https://doi.org/10.1016/j.anbehav.2021.01.024>

### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the KNAW public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the KNAW public portal.

### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

### **E-mail address:**

[pure@knav.nl](mailto:pure@knav.nl)

# Animal Behaviour

## Motivation, accuracy and feedback through experience explain innovative problem solving and its repeatability --Manuscript Draft--

<b>Manuscript Number:</b>	ANBEH-D-20-00198R2
<b>Article Type:</b>	UK Research paper
<b>Keywords:</b>	accuracy; cognitive repeatability; inhibitory control; innovation; motivation; Parus major; personality; problem solving; pseudo-repeatability
<b>Corresponding Author:</b>	Amy C. Cooke University College Cork National University of Ireland Cork, Cork IRELAND
<b>First Author:</b>	Amy C. Cooke
<b>Order of Authors:</b>	Amy C. Cooke Gabrielle L. Davidson Kees Van Oers John L. Quinn
<b>Abstract:</b>	<p>Adapting to environmental change is a major challenge faced by animals and the role of individual behavioural differences in facilitating this process is currently the focus of much research. Innovation, the generation of a novel behaviour or a known behaviour used in a novel context, is one form of behaviour that enables animals to respond to change. By deciphering the mechanisms underlying innovativeness, especially those that explain consistent differences between individuals, we can further understand the consequences of this behavioural variation. We tested whether motivation, experience, inhibitory control, and personality were linked to different stages of sequential innovative problem-solving performance among great tits (<i>Parus major</i>), and of their overall innovativeness across tasks. We gave animals originating from lines bi-directionally selected for fast or slow early exploratory behaviour a multi-access problem-solving device with three different access points, requiring diverse motor skills and behavioural flexibility to solve sequentially over trials. Food deprived, highly motivated birds had shorter latency to touch the device, were more likely to solve within a trial, and solved a greater diversity of problems than their less motivated counterparts. Solving success increased with accuracy (proportion of touches to functional components of the device compared to all touches to the device per trial) when interacting with the device, and with previous experience. Personality selection lines and inhibitory control had little effect. Repeatability analysis showed that between individual differences in problem solving performance were explained by: i) pseudorepeatable effects (upwards bias) linked to hunger-induced motivation; ii) by repeatable differences in accuracy when interacting with devices, and iii) by a feedback loop caused by experience gained over successive trials. Our results highlight the challenges of characterising consistent individual variation in behaviour generally and supports the idea that complex sources of variation play an important role in problem solving performance.</p>

1 Adapting to environmental change is a major challenge faced by animals and the role of  
2 individual behavioural differences in facilitating this process is currently the focus of much  
3 research. Innovation, the generation of a novel behaviour or a known behaviour used in a novel  
4 context, is one form of behaviour that enables animals to respond to change. By deciphering  
5 the mechanisms underlying innovativeness, especially those that explain consistent differences  
6 between individuals, we can further understand the consequences of this behavioural variation.  
7 We tested whether motivation, experience, inhibitory control, and personality were linked to  
8 different stages of sequential innovative problem-solving performance among great tits (*Parus*  
9 *major*), and of their overall innovativeness across tasks. We gave animals originating from  
10 lines bi-directionally selected for fast or slow early exploratory behaviour a multi-access  
11 problem-solving device with three different access points, requiring diverse motor skills and  
12 behavioural flexibility to solve sequentially over trials. Food deprived, highly motivated birds  
13 had shorter latency to touch the device, were more likely to solve within a trial, and solved a  
14 greater diversity of problems than their less motivated counterparts. Solving success increased  
15 with accuracy (proportion of touches to functional components of the device compared to all  
16 touches to the device per trial) when interacting with the device, and with previous experience.  
17 Personality selection lines and inhibitory control had little effect. Repeatability analysis  
18 showed that between individual differences in problem solving performance were explained  
19 by: i) pseudorepeatable effects (upwards bias) linked to hunger-induced motivation; ii) by  
20 repeatable differences in accuracy when interacting with devices, and iii) by a feedback loop  
21 caused by experience gained over successive trials. Our results highlight the challenges of  
22 characterising consistent individual variation in behaviour generally and supports the idea that  
23 complex sources of variation play an important role in problem solving performance.

24 **Keywords:** accuracy, cognitive repeatability, inhibitory control, innovation, motivation, *Parus*  
25 *major*, personality, problem-solving, pseudo-repeatability.

26 Acquiring resources in changing environments is a major challenge faced by animals and a key  
27 determinant of fitness. Innovation, the generation of a novel behaviour or a known behaviour  
28 used in a novel context, most commonly achieved through a problem solving process, is one  
29 mechanism that a wide range of animals use to meet this challenge (Seed & Mayer, 2017).  
30 Comparative analysis has provided evidence for selection acting on innovativeness across  
31 species, such as the need to find food, adapting to urban environments, and seasonal changes  
32 (Daniels et al., 2019; Lefebvre et al., 2004; S. Reader, 2003; Simon M. Reader & Laland, 2002;  
33 Sol, Lefebvre, et al., 2005; Webster & Lefebvre, 2001). Furthermore there is growing evidence  
34 of a correlation between innovativeness and fitness within populations (Cauchard et al., 2013;  
35 Cole et al., 2012; Preiszner et al., 2017), and for the role of innovation among invasive or  
36 urbanised species (Daniels et al., 2019; Griffin et al., 2014; Griffin & Diquelou, 2015), enabling  
37 them to make use of novel resources. Although the underlying proximate causes of individual  
38 variation in innovativeness are diverse - for example, infection by parasites (Dunn et al., 2011),  
39 social factors (Thornton & Samson, 2012) and natal environment effects (Kotrschal &  
40 Taborsky, 2010) - repeatability analyses suggest differences among individuals are consistent,  
41 pointing to intrinsic, potentially additive genetic, sources of variation (Cauchoix M. et al.,  
42 2018; Cole et al., 2011; Morand-Ferron et al., 2011). One of the major challenges in the field  
43 is that innovativeness is a composite trait driven by a range of disparate behavioural processes  
44 that selection may act on independently and that may explain consistent differences in  
45 performance among individuals. These processes include cognition and motivation, as well as  
46 personality traits like exploration, persistence and neophobia (Griffin & Guez, 2014; Lermite  
47 et al., 2017; Seed & Call, 2010; Taylor et al., 2009). Thus, a key objective is to determine which  
48 processes drive innovativeness and explain the consistent individual differences observed.

49 Innovativeness correlates with relative brain size across species (Overington et al., 2009),  
50 suggesting an important role for cognitive mechanisms (Benson-Amram et al., 2016; Lefebvre

51 et al., 1997; Simon M. Reader & Laland, 2002; Sol, Duncan, et al., 2005). A number of  
52 cognitive mechanisms have been proposed to underlie innovative problem-solving, including  
53 causal reasoning, insight, associative learning and inhibitory control (Barrett et al., 2018;  
54 Benson-Amram & Holekamp, 2012). Additionally, the ability to draw on previous experience  
55 aids individuals in reaching a solution (Cauchard et al., 2013; Ebel & Call, 2018; Griffin et al.,  
56 2014; Sol et al., 2012), as well as ensuring the new behaviour becomes established in an  
57 individual's repertoire (Fragaszy et al., 2013). Finally, in situations where opportunities for  
58 innovation may be common, for example among invasive or urbanised species (Daniels et al.,  
59 2019; Griffin et al., 2014; Griffin & Diquelou, 2015), the ability to innovate frequently may be  
60 essential but dependant on inhibitory control (Daniels et al., 2019), whereby individuals must  
61 inhibit a previously rewarding behaviour that is no longer rewarding, in order to innovate  
62 further (Pecora et al., 2017). However, cognitive processes alone cannot explain why some  
63 individuals innovate more compared to others (van Horik & Madden, 2016), and especially  
64 when other behavioural traits such as exploration and persistence also lead to innovation  
65 (Daniels et al., 2019; Ebel & Call, 2018; Overington et al., 2011).

66 Personality, defined as within individual behavioural consistency across time and contexts  
67 (Réale et al., 2007), provides a framework for exploring constraints on behavioural plasticity  
68 (Dall et al., 2004) and individual problem-solving performance (Hopper et al., 2014; Morton  
69 et al., 2013). Personality traits have attracted particular attention because they predict  
70 individual variation in a wide range of behavioural traits (Aplin et al., 2014; Cole & Quinn,  
71 2012, 2014). Studies in the wild (Dingemanse et al., 2002; Highcock & Carter, 2014) and  
72 laboratory (David et al., 2012; van Oers & Naguib, 2013), show that the personality trait 'early-  
73 life exploratory behaviour', more specifically in this case, repeatable differences in the reaction  
74 to both a novel environment and objects (Drent et al., 2003), can influence how individuals  
75 retrieve information from their environment (Smit & van Oers, 2019), how quickly they solve

76 problems (Hopper et al., 2014), and the degree of behavioural flexibility shown (Coppens et  
77 al., 2010). In particular, fast-exploring (hereafter fast) individuals may be quicker to interact  
78 with or solve tasks (Benson-Amram & Holekamp, 2012; Trompf & Brown, 2014) but show  
79 less behavioural plasticity (Amy et al., 2012; Jolles et al., 2019; Logan, 2016c). Slow-exploring  
80 (hereafter slow) individuals tend to be the opposite (Coppens et al., 2010; Ducatez et al., 2015;  
81 Johnson-Ulrich et al., 2018; Sol et al., 2012; Zandberg et al., 2017). Additionally, neophobia,  
82 the fear of novel food, objects or places (R. S. Greenberg & Mettke-Hofmann, 2001), can  
83 constrain both the latency to approach a novel object and engagement in tasks. For example,  
84 individual hyenas that show greater persistence, activity, or lower neophobia, were faster to  
85 solve a problem. (Johnson-Ulrich et al., 2018). However, the evolutionary significance of links  
86 between innovation and personality traits, as defined in Reale et al. (2007), is often unclear  
87 because the genetic basis for the personality variation is usually unknown (Cole et al., 2011),  
88 except in those few cases where personality selective breeding lines have been used (Drent et  
89 al., 2003; van Oers et al., 2004, 2005). Moreover, the role of other personality traits at different  
90 stages of innovative problem-solving (e.g. interacting with a problem, solving a problem, and  
91 ceasing to perform outdated solutions), and its interactions with other factors such as stress and  
92 motivation remains largely unexplored. Individual differences may be especially pronounced  
93 under stress (Suomi, 2004), but this has scarcely been tested. Note that although all behavioural  
94 variation can be defined as personality in a statistical sense (e.g Dingemanse and Dochterman  
95 2013), here we follow Reale et al. (2007) who focus on five kinds of behavioural traits,  
96 including exploration behaviour, that inherently capture variation in many other behavioural  
97 traits.

98 Motivation is expected to be an important driver of innovative behaviour (Laland & Reader,  
99 1999; Sol et al., 2012) and to affect all stages of innovation. The “necessity drives innovation”  
100 hypothesis states that innovative behaviours commonly occur when individuals are in need

101 (S.M. Reader & Laland, 2003), i.e. when they are motivated (Laland & Reader, 1999). For  
102 example, subdominant or juvenile individuals are often assumed to be more likely to innovate  
103 because they are less competitive when foraging (Morand-Ferron et al., 2011; Thornton &  
104 Samson, 2012). The rarely tested assumption in these studies is that hunger acts as the  
105 motivating factor driving innovation. In animal behaviour studies, food deprivation is  
106 commonly applied to ensure trial participation (Birch, 1945; Overington et al., 2011; Sol et al.,  
107 2012), or when attempting to control for confounding effects of motivation (Ebel & Call, 2018;  
108 van Horik & Madden, 2016). However, the extent to which motivation may influence  
109 innovative problem-solving behaviour at an individual level has scarcely been examined  
110 explicitly (Griffin and Guez, 2014).

111 Here we explore behavioural processes that are predicted to cause variation during sequential  
112 innovative problem-solving, using second and third generation birds selected for personality.  
113 Selection lines are a powerful means to investigate inherent effects of personality on problem  
114 solving performance as opposed to simple phenotype-phenotype correlations. We used a device  
115 that incorporated three different extractive foraging access points to provide a more complete  
116 measure of individual ability. The solutions relied on different motor skills, thus limiting the  
117 effects of individual motor skill bias, and previous motor skill experience carrying over to  
118 solving new access points. We examined variation in three different behavioural assays  
119 involved in innovative problem-solving: i) latency to touch the novel apparatus; ii) accuracy  
120 when interacting with any access point on the device; iii) problem-solving success within each  
121 trial, before finally examining iv) the individual's overall innovativeness (the number of  
122 different access points solved at least once across all trials). We considered a range of potential  
123 explanatory factors for these different behavioural facets, including extrinsic motivation  
124 (hunger state, the only experimentally manipulated factor), inhibitory control, previous  
125 experience, and personality (fast/slow selection lines). In line with the theory that individual

126 differences may be more pronounced under stress (Suomi, 2004), we investigated the  
127 interaction between motivation and personality, assuming that birds in the high motivation  
128 (food deprived) treatment were more stressed than those in the low motivation treatment.  
129 Finally, to determine whether individual differences were consistent, we estimated  
130 repeatability for i)-iii) and examined whether controlling for fixed effects modified our  
131 estimates of repeatability. Repeatability sets the upper limit of heritability and is fundamental  
132 in studies on the evolutionary ecology of innovation and behaviour generally. Although  
133 uncontrolled confounding effects can potentially lead to an underestimate of repeatability,  
134 more commonly they lead to overestimates (pseudo-repeatability) and sometimes explain  
135 repeatability entirely (Catry et al., 1999; Dingemanse & Dochtermann, 2013; Westneat et al.,  
136 2011).

137 We predicted that: 1) Birds in the high motivation treatment group would have a reduced  
138 latency to touch the device, increased accuracy (i.e. a high proportion of interactions with  
139 functional components, rather than non-functional components of the device), would be more  
140 likely to solve, and solve a greater overall number of different access points; 2) Fast explorers  
141 would have a shorter latency to touch the device and lower accuracy when interacting with the  
142 device than slow explorers, but they may have a higher likelihood of solving due to higher  
143 exploration of the device; 3) Previous experience would enable innovation, by causing a  
144 decrease in latency to touch the device, and an increase in accuracy with interacting with it and  
145 likelihood of solving; 4) Likelihood of solving in a trial would increase with accuracy (i.e. with  
146 higher interactions with functional components); and 5) Birds with higher inhibition ability  
147 would be more likely to adjust their behaviour to solve multiple access points.

148 MATERIALS AND METHODS



149 All experiments were carried out at the Netherlands Institute of Ecology (NIOO-KNAW), on  
150 36 captive bred great tits (*Parus major*). All birds included in the study were adult (two years  
151 or older). 17 birds were not related to each other, and 5 had one sibling and 14 shared more  
152 than one sibling; we assume relatedness among individuals had no bearing on the results. They  
153 were housed individually in standard cages (0.9 x 0.5 m and 0.5 m high) containing three  
154 perches and a water bath. Birds were in auditory contact but were visually isolated to prevent  
155 social learning. All birds had ad libitum access to water and a maintenance diet (ground beef  
156 heart, commercial egg food, fruit and calcium) unless otherwise stated. One bird did not  
157 participate in any of the experiments, it was excluded from any analysis due to the absence of  
158 any observations.

### 159 *Personality*

160 Birds came from the second and third generation of bi-directionally phenotypically selected  
161 great tits, based on personality for “fast exploration” (fast;  $N = 18$ ) and “slow exploration”  
162 (slow;  $N = 18$ ). The measure of “exploration” used during the selection process, was a  
163 combination of two novel object tests where the latency to touch a novel object was recorded,  
164 e.g. a pink panther toy or a AA battery taped to a wooden stick, and one novel environment  
165 test, where birds were released into a room and the latency to land on the fourth (out of five)  
166 artificial trees was recorded (for further details on selection and personality lines see Drent et  
167 al., 2003). The birds in the final selection lines used here underwent these same assays after  
168 fledging to confirm their personality type. As the specific aim of this study was to investigate  
169 the effects of artificially selected personality lines on problem-solving and because the bird  
170 behaviour matched their selected personality type, we analysed personality according to their  
171 selection history only (i.e. fast or slow selection lines).

### 172 *Motivation*

173 Individuals were randomly assigned to one of two motivation treatment groups for the duration  
174 of the experiment based on hunger state. The low motivation group consisted of sated  
175 individuals, given full access to maintenance diet up to the start of the trial. Additionally to  
176 ensure they were sated, they were provided with and invariably ate all three wax moth larvae  
177 30 minutes before trials began. The high motivation group consisted of food deprived birds,  
178 who had all sources of food removed from their cage for 1 hour before the trial (Hämäläinen  
179 et al., 2019). For welfare purposes, all birds had access to water during the trials. Motivation  
180 treatment was spread across the selection lines in four categories; high motivation, fast (*female*  
181 *N* = 4 and male *N* = 5), high motivation, slow (female *N* = 5 and male *N* = 4), low motivation,  
182 fast (female *N* = 5 and male *N* = 4), low motivation, slow (female *N* = 5 and male *N* = 3).

### 183 *Lever-pulling propensity*

184 All trials described here and in the following section were carried out in their individual home  
185 cages, under natural winter diurnal light cycles. To establish whether the birds had a pre-  
186 existing tendency to lever-pull (von Bayern et al., 2009), and because some birds may have  
187 had previous experience lever pulling in old experiments while others had not, we measured  
188 lever pulling propensity prior to testing them on the multi-access device. We presented all birds  
189 with an opaque PVC rectangular tube containing a lever-supported platform with half a wax  
190 worm (Zandberg et al., 2017). We used an opaque device to test whether birds had a propensity  
191 to pull a stick, independent of a visual food reward cue, because the previous device that had  
192 been used was also opaque (Zandberg et al., 2017), and because we did not want the birds to  
193 have experience with the main innovation test device beforehand. All birds were given up to  
194 four trials (30minutes per trial) to obtain the food reward, by pulling the lever horizontally  
195 causing the platform and reward to drop. Individuals who solved this opaque task at least once  
196 were classified, in the main analysis on the multi-access problem solving task described below,

197 as having previous experience with solving the lever pulling task. All birds progressed to the  
198 multi access task irrespective of their performance in this opaque device (see figure 1).

### 199 *Multi Access Problem-solving Task*

200 Birds were presented with a multi access problem-solving apparatus (figure 2) with three  
201 distinct solutions that required different motor skills (see below), to obtain a preferred food  
202 reward (a wax moth larvae). The apparatus was an upright Perspex cylinder (height = 16cm,  
203 diameter = 5cm), with a platform holding the food reward. The platform was supported by a  
204 lever, which when pulled from the outside of the device caused the platform to drop, releasing  
205 the food reward below the device (solution 1). A second possible solution was to move a door  
206 that could be pushed left or right, to gain access to the food reward on the platform (solution  
207 2). A third possible solution was to pull a string from the top of the device, which was attached  
208 to a second worm (solution 3). Each of these access points involved different motor action(s)  
209 including pulling (solution 1), pushing (solution 2) and coordinating both grasping and pulling  
210 (solution 3).

211 Experiments were scheduled evenly across mornings and afternoons for both treatments and  
212 personality lines. In each trial, subjects were presented with the device and given 30 minutes  
213 to solve any of the access points. Birds were given two back to back trials per day, without  
214 being fed between trials. Following their second trial, their maintenance diet was returned until  
215 testing the following day. The experiment ended when they had solved all three access points  
216 three times, or until they had failed to solve over three consecutive trials (total number of trials  
217 ranged between 3 and 13). Once an individual solved the same access point across three  
218 separate trials, that access point (door, lever, or string) was fused mimicking natural depletion  
219 of that food source, which meant that solving that particular access point was no longer  
220 possible, though remained present and visible. We allowed birds to solve each access point

221 three times to increase the chance that the behaviour became fixed in their repertoire. In order  
222 to solve a novel solution, they would need to behave flexibly, which we predicted would be  
223 guided by inhibitory control. Great tits from selection lines in this facility readily participate in  
224 experiments, so we assumed the three trials was sufficient to allow them overcome any  
225 neophobic response. All trials were recorded using a Panasonic HC-V250EB-K camera  
226 mounted on a tripod, covered in camouflage tape and positioned 1m from cage. Videos were  
227 analysed using Behavioural Observation Research Interactive Software - BORIS (Friard &  
228 Gamba, 2016). Observers were blind to the personality assigned to the birds, but were aware  
229 of the motivation treatment group. Ten per cent of videos were coded by a second person. Inter-  
230 rater reliability was assessed using a Kendall's Tau correlation test for agreement on the  
231 following measures; total number of touches to the device per trial  $P < 0.001$ ; touches to  
232 functional access points on the device per trial,  $P < 0.01$ , touches to anything other than  
233 functional access points on the device per trial  $P < 0.005$ .

#### 234 *Inhibition task*

235 To generate an independent estimate of each individual's motor inhibition, we used a classical  
236 detour reaching task (Beran, 2015; Boogert et al., 2011; Rothbart et al., 1985), which tests to  
237 what extent the birds could control the prepotent response of pecking straight towards a food  
238 reward visible within a transparent Perspex tube. To pass the test, birds had to obtain the reward  
239 by accessing it through the opening on the side (Thorndike, 1911). The detour task was  
240 performed on a subset of twenty birds, prior to the problem-solving task (number of days  
241 between end of detour reach task and first test day on multi-access device; mean  $\pm$  s.e. =  $11 \pm$   
242 0.46, min = 8, max = 12) to control for carry-over experience with the transparent Perspex.  
243 Birds were not food deprived before this task. There were three phases to this task: habituation,  
244 training, and test phases. Birds participated in one phase per day, with progression through

245 phases occurring over consecutive days (duration of testing (days); mean  $\pm$  s.e. = 1.64  $\pm$  0.18,  
246 min = 1 , max = 4). In the habituation and training stages, the Perspex tube was opaque (covered  
247 with black tape). To familiarise the birds with the apparatus, a wax worm larva was placed at  
248 the opening edge of the tube. Birds passed the habituation phase when they had eaten the  
249 reward three consecutive times. During the training phase, individuals had to obtain the food  
250 reward located in the centre of the opaque tube without touching any other part of the device.  
251 Training was completed when this was successfully demonstrated during four out of five  
252 consecutive trials, ensuring the birds had the motor skills and experience necessary to move  
253 around the tube to successfully obtain the worm. During the test phase, the food reward was  
254 placed in the centre of a transparent tube. Birds had to remove the food reward without pecking  
255 on any other part of the device to complete the trial successfully. Inhibitory control scores were  
256 quantified as the number of trials it took individuals to complete four out of five consecutive  
257 trials correctly. All trials were a maximum of three minutes each, and observed remotely by  
258 live streaming to a mobile phone using a WIFI enabled SJCAM SJ4000 camera.

#### 259 *Ethical Note*

260 We performed the experiment in accordance with the Association for the Study of Animal  
261 Behaviour guidelines. All experiments were approved by an ethical committee (DEC-KNAW  
262 licence no. NIOO 14.12 to KVO) and daily health checks were carried out to ensure the birds'  
263 welfare. Birds returned to the stock population after the behavioural experiments.

#### 264 *Statistical analysis*

265 We tested whether multiple factors influenced different behaviours ((i) latency to touch, (ii)  
266 accuracy, (iii) likelihood of solving, (iv) innovativeness) at different stages of sequential  
267 innovative problem-solving performance. Separate analyses were conducted using RStudio  
268 (RStudio Team, 2019) on each of the four stages above (i-iv), and we repeated these models

269 on the subset of birds ( $n = 22$ ) who completed the inhibition task. For touch latency and  
270 accuracy, we conducted general linear mixed models (GLMMs) using the nlme package (J  
271 Pinheiro et al., 2019) fit with a normal distribution; for likelihood of solving we ran a GLMM  
272 using the lme4 package (Bates et al., 2015) fit with a binomial distribution; and for  
273 innovativeness we ran a general linear model (GLM) with poisson distribution (see table A1  
274 for a full list of variables and their definitions). In line with Whittingham et al., (2006), we  
275 retained all variables of biological significance in the initial models to test specific hypotheses.  
276 For model selection, we used Akaike's information criterion (AIC) to measure goodness of fit  
277 (reported in table legends) and likelihood-ratio tests to determine which model explained more  
278 variance. We compared full models (with the interaction between motivation and personality)  
279 to null models, and then compared full models to reduced models (i.e. without the interaction  
280 between motivation and personality). We dropped the interaction term from the model if the  
281 likelihood-ratio test was non-significant ( $\alpha = 0.05$ ). To confirm that this hypothesis testing  
282 approach did not lead to a Type 2 error due to overfitting, we further reduced each model to  
283 the minimum adequate model using backwards reduction (see tables A2-A5). We checked that  
284 all models met assumptions (homogeneity, normality of residuals and collinearity of  
285 explanatory variables) using the DHARMA package in R (Hartig, 2020). We calculated  
286 confidence intervals (C.I.) for the random factor and residuals in each model using the package  
287 nlme in R (José Pinheiro & Bates, 2006). In the legend of each table, we report marginal  $R^2$   
288 (defined as the proportion of variance in the dependent variable that is explained by the fixed  
289 factors only), and conditional  $R^2$  (defined as the proportion of variance in the dependent  
290 variable that is explained by the fixed and random factors), or pseudo  $R^2$  (the marginal  $R^2$  of a  
291 Poisson GLM, which does not include a random factors). The full models tested in the analyses  
292 were as follows:

293 i) Latency to touch the device, was log-transformed to fit a Gaussian distribution (total trials =  
294 226). The following fixed effects were included in our model as potential influences:  
295 motivation (low or high), trial number, selected personality lines (fast or slow exploring),  
296 previous experience of solving any functioning access point including the opaque device (no  
297 or yes), and sex (male or female). Individual bird identity was included as a random effect to  
298 control for repeated measures and to test repeatability of individual differences.

299 ii) Accuracy was defined as the number of touches to a functioning access point divided by the  
300 total number of touches to any part of the device per trial. Fixed effects included interaction  
301 rate (total number of touches to any part of the device  $\text{min}^{-1} \text{trial}^{-1}$ ), motivation group, trial  
302 number, selected personality lines, previous experience of any functioning (but not fused)  
303 access point (including previous experience of lever pulling propensity on opaque device),  
304 fused trial – where any of the solutions fused and therefore unavailable, as a fused access point  
305 may decrease accuracy, and sex. Individual bird identity was included as a random effect as  
306 subjects completed multiple trials.

307 iii) To test which factors predicted solving within each trial (binary;  $n = 224$ ), we included the  
308 following fixed effects: accuracy, previous experience, motivation group, personality, sex, trial  
309 number and fused trial. Individual bird identity was included as a random term. To avoid over  
310 parameterisation of the model, we did not include latency to touch in this particular analysis  
311 (but see analysis on number of different solves).

312 iv) We tested which factors affected innovativeness defined as the number of different access  
313 points solved by an individual, ( $n = 35$ ). Birds solved either 0, 1, 2, or 3 different access points.  
314 We included the following explanatory variables: hunger, personality, sex, latency to touch the  
315 device in the first trial only and inhibitory control. As this analysis was conducted on the

316 number of different access points of solved across all trials, we do not include variables that  
317 are trial specific (i.e. previous experience and accuracy).

318 Finally, we determined individual repeatability of the response variables in each of the first  
319 three questions above (latency to touch the device, accuracy and solving within a trial), using  
320 the rptR package, estimating repeatability (intra-class correlation) and confidence intervals  
321 (C.I.) from Gaussian, binary, proportion and Poisson data (Stoffel et al., 2017). We report  
322 unadjusted and adjusted repeatability, to encompass repeatability before and after controlling  
323 for influential fixed effects (Cauchoix M. et al., 2018). Unadjusted repeatability measures the  
324 between-individual variation in a given behaviour, while adjusted repeatability controls for  
325 fixed effects that could influence individual behaviour, either because they explain between or  
326 within individual components of variation. For both adjusted and unadjusted repeatability, we  
327 included individual identity as a random effect.

## 328 RESULTS

### 329 *Latency to touch the multi-access device*

330 Latency to touch the multi-access device decreased over consecutive trials ( $\beta \pm \text{s.e.} = -0.09 \pm$   
331  $0.03$ ,  $t = -3.36$ ,  $P = 0.001$ ; table 1, supplementary material figure A1). The high motivation  
332 group took less time to touch the device than the low motivation group ( $\beta \pm \text{s.e.} = 1.17 \pm 0.30$ ,  
333  $t = 3.84$ ,  $P < 0.001$ ). Latency to touch the device did not differ between the personality selection  
334 lines ( $\beta \pm \text{s.e.} = 0.20 \pm 0.30$ ,  $t = 0.67$ ,  $P = 0.507$ ). There was a non-significant trend for sex,  
335 suggesting that males took less time to touch the device compared to females ( $\beta \pm \text{s.e.} = -0.53$   
336  $\pm 0.30$ ,  $t = -1.76$ ,  $P = 0.089$ ). The variance of the random effect (individual bird identity) was  
337  $0.73$  and the residual was  $1.07$  (95% C.I. [0.97, 1.18]). There was no effect of previous  
338 experience. The interaction between motivation and personality was not significant ( $\beta \pm \text{s.e.} =$



339  $-0.47 \pm 0.60$ ,  $t = -0.78$ ,  $p = 0.44$ ). Inhibitory score had no effect on latency to touch the device  
340 (see supplementary material table A6).

### 341 *Accuracy*

342 Birds were more accurate if they had previous experience solving any functioning access point,  
343 including solving the opaque device before the main experiment ( $\beta \pm \text{s.e.} = -0.35 \pm 0.04$ ,  $t =$   
344  $8.20$ ,  $P < 0.001$ ; table 2 and figure 3). There was a non-significant trend for slow birds being  
345 more accurate than fast birds ( $\beta \pm \text{s.e.} = 0.13 \pm 0.07$ ,  $t = 1.80$ ,  $P = 0.08$ ). Birds tended to be less  
346 accurate in trials where there was a fused access point ( $\beta \pm \text{s.e.} = -0.11 \pm 0.07$ ,  $t = -1.70$ ,  $P =$   
347  $0.09$ ). The variance of the random effect (individual bird identity) was 0.17 and residual was  
348 0.24 (95% C.I. = 0.22-0.27). There was no effect of motivation group, interaction rate, sex or  
349 trial number on accuracy. The interaction between motivation and personality was not  
350 significant ( $\beta \pm \text{s.e.} = -0.06 \pm 0.15$ ,  $t = -0.39$ ,  $p = 0.70$ ). Inhibition was unrelated to accuracy  
351 (see supplementary table A7).

### 352 *Solving within a trial*

353 Nineteen of the 35 birds pulled the lever on the opaque device. Of the 35 birds that participated  
354 in the multi-access task, 12 birds solved one access point, four birds solved two access points,  
355 seven birds solved all three access points, and 12 birds did not solve at all (figure 4). Three  
356 birds solved three different access points over three consecutive trials while the device was  
357 fully operational (all access points functioning). One bird solved two access points in one trial,  
358 solving the string and then the lever in their fourth trial. We include both solves as separate  
359 observations in our analysis. The lever was solved by 23 different birds, the door was solved  
360 by 10 birds and the string was solved by 7 birds, while 12 birds did not solve any access point.

361 Food deprived birds were more likely to solve ( $\beta \pm \text{s.e.} = -1.79 \pm 0.50, z = -3.57, P < 0.001$ ;  
362 table 3). Higher accuracy ( $\beta \pm \text{s.e.} = 3.79 \pm 0.78, z = 4.90, P = 0.008$ ) and previous experience  
363 ( $\beta \pm \text{s.e.} = 1.26 \pm 0.47, z = 2.66, P = 0.008$ ) also predicted solving likelihood within a trial. The  
364 variance of the individual identity random effect was 0.13 and residual was 0.37 (95% C.I.  
365 [0.41, 0.33]). There was no effect of personality, sex, trial number, whether it was a fused trial  
366 or not, total number of touches to device per trial or inhibition (see supplementary table A8).  
367 The interaction between motivation and personality was not significant ( $\beta \pm \text{s.e.} = -0.46 \pm 0.90$ ,  
368  $t = -0.51, p = 0.61$ ). Follow up post hoc analysis, using a Fisher's exact test, revealed a  
369 correlation trend between lever-pulling on the opaque and multi-access device ( $P = 0.07$ ).  
370 Further analysis, investigating the order in which the multi-access device was solved, using a  
371 Fisher's exact test, showed that the lever was more likely to be solved first ( $P < 0.001$ ), while  
372 there was no difference between the string or door (see figure A2).

### 373 *Innovativeness: number of access points solved*

374 Highly motivated birds solved a greater number of novel access points than little motivated  
375 birds ( $\beta \pm \text{s.e.} = -1.04 \pm 0.37, z = -2.80, P = 0.005$ ; table 4 and figure 5). There was no effect of  
376 personality, sex, latency to touch the device in the first trial only (see table 4) or inhibition (see  
377 supplementary table A9). The interaction between personality and motivation was non-  
378 significant ( $\beta \pm \text{s.e.} = -0.86 \pm 0.79, z = -1.09, p = 0.28$ ).

### 379 *Repeatability*

380 Latency to touch the device was repeatable but repeatability decreased when adjusted for  
381 significant fixed effects (unadjusted  $R = 0.48$ , C.I. = 0.30, 0.61,  $P < 0.001$ ; adjusted  $R = 0.32$ ,  
382 C.I. = 0.15, 0.45,  $P < 0.001$ ; table 5). Accuracy was also repeatable but increased when adjusted  
383 for significant effects (unadjusted  $R = 0.23$ , C.I. = 0.08, 0.38,  $P = 0.001$ ; adjusted  $R = 0.35$ ,  
384 C.I. = 0.17, 0.52,  $P < 0.001$ ). Solving performance within a trial was also repeatable but

385 repeatability disappeared entirely when adjusted for all significant fixed effects (unadjusted R  
386 = 0.31, C.I. = 0.1, 0.46,  $P < 0.001$ ; adjusted R = 0.03, C.I. = 0, 0.17,  $P = 0.40$ ). To further  
387 investigate which factors were reducing the individual repeatability between the unadjusted  
388 and adjusted R-values for solving within a trial, we removed each fixed effect individually and  
389 re-ran the repeatability model (see table 5). Adjusted repeatability changed only when a  
390 significant fixed effect was excluded. Adjusted repeatability without accuracy was significant  
391 (adjusted R = 0.18, C.I. = 0, 0.33,  $P = 0.044$ ) and without motivation (R = 0.21, C.I. = 0, 0.35,  
392  $P = 0.017$ ) while adjusted repeatability without previous experience only trends towards  
393 significant (R = 0.14, C.I. = 0.0, 0.29,  $P = 0.076$ ). There was no change in adjusted repeatability  
394 for any factor that did not affect solving.

## 395 DISCUSSION

396 Our study sought to explore factors that drive individual variation and repeatability at various  
397 stages of innovative problem-solving performance (summary figure 6). We show that hunger-  
398 induced motivation affected multiple problem-solving stages, that previous experience  
399 influenced accuracy, and that hunger, accuracy, and previous experience influenced problem  
400 solving success. Personality and inhibitory control had little effect. Solvers of the opaque lever  
401 pulling device tended to solve the lever on the multi-access device. Furthermore, birds were  
402 more likely to solve the lever first, but showed no preference between the door and string. All  
403 traits were significantly repeatable; however, the repeatability of problem solving was  
404 explained entirely by motivation, accuracy and experience.

### 405 *Motivation drives innovation*

406 Although motivation is often viewed as a confounding variable, if considered at all when  
407 examining mechanisms underlying problem-solving tasks (reviewed in Griffin and Guez  
408 2014), it also underpins the “necessity drives innovation” hypothesis (S.M. Reader & Laland,

409 2003). In support of this hypothesis, motivation was the major driver of an individual's latency  
410 to touch the device, to solve the same access point repeatedly, and to innovate multiple times  
411 in our experimental setup. Previous studies report that task engagement increased with  
412 increased food deprivation, thus facilitating problem-solving (Griffin et al., 2014; Sol et al.,  
413 2012), but motivation itself did not predict problem-solving (Griffin & Guez, 2014; van Horik  
414 & Madden, 2016). Likewise, the relationship between problem-solving and motivation, as  
415 measured by body weight or body condition is inconclusive - at times an effect is present  
416 (Laland & Reader, 1999; Mateos-Gonzalez et al., 2011) and other times not (Cole et al., 2011;  
417 Thornton & Samson, 2012)but see Griffin Guez, 2014 for full review); variability that may be  
418 in part due to differences in how motivation is defined and how problem-solving is measured.  
419 Our results emphasise the importance of controlling for motivation, and highlights the  
420 importance of standardising the length of time animals are food deprived in captive  
421 experiments, as well as acknowledging this as a weakness of cognitive experiments conducted  
422 in the wild. Nevertheless controlling for motivational effects is unlikely to be so  
423 straightforward (Auersperg et al., 2012; Griffin & Guez, 2014; Morand- Ferron et al., 2016;  
424 Morand-Ferron & Quinn, 2011), not least because whether food deprivation removes, or just  
425 changes, individual variation remains unclear.

#### 426 *Personality*

427 Considerable evidence suggests that personality traits defined by Reale et al., (2007) influence  
428 problem-solving performance among individuals (R. Greenberg, 2003; Johnson-Ulrich et al.,  
429 2018; Sol et al., 2011). However, in our study, personality selection lines with known genetic  
430 provenance for object neophobia and novel environment exploration did not predict latency to  
431 touch the device; nor did it predict problem-solving behaviour, in terms of success within trials,  
432 or the number of different innovations reached. We predicted the effects of hunger-induced

433 motivation could mask effects of artificially selected personality lines on problem solving  
434 behaviour, but the interaction between motivation and personality had no effect on any  
435 problem-solving measure, suggesting that our ability to detect the effect of personality on an  
436 individual's capacity to problem solve was not confounded by motivation or vice versa.  
437 Furthermore, while there was a non-significant tendency for slow birds to be more accurate,  
438 this did not translate into higher likelihood of solving and innovativeness for slow birds.  
439 Previous work in this same population, using a lever pulling task, also found no link between  
440 personality and innovative problem solving performance (Zandberg et al., 2017). The absence  
441 of an effect of personality on problem solving performance in that study, and here, could be  
442 influenced by the composite nature of "exploration" used in our selection lines (Verbeek et al.,  
443 1994). Moreover, latency to touch, which may be considered a measure of neophobia, may  
444 have been confounded with associative learning when considering latency to touch across  
445 multiple trials. Nevertheless, our results emphasise the challenge of examining links between  
446 personality traits and innovative problem solving, not least because of the inherently composite  
447 nature of both behaviours.

#### 448 *Inhibitory control*

449 Inhibitory control is an integral part of behavioural flexibility (MacLean et al., 2014; Manrique  
450 et al., 2013), both of which are beneficial for problem-solving, allowing animals to overcome  
451 outdated information. Contrary to our predictions, individuals that exhibited high inhibitory  
452 control were no more likely to generate a novel solution to the task than those with low  
453 inhibitory control, even when the reward contingencies changed (i.e. when an access point was  
454 fused) - a time when behavioural flexibility is required. This lack of correlation may be because  
455 changing ones' behaviour is necessary but not sufficient to solve a problem, (Logan, 2016a,  
456 2016b). Moreover, the validity of the detour reaching task as a test for inhibitory control

457 remains under debate because performance does not necessarily correlate with other tasks that  
458 aim to measure inhibitory control, or because previous experience of transparency, and  
459 persistence may influence performance (Kabadayi et al., 2018; van Horik et al., 2018). Neither  
460 Johnson-Ulrich (2018) nor Daniels et al (2019) found a correlation between problem-solving  
461 and inhibitory control even when inhibitory control was measured using an alternative  
462 paradigm to the detour task. Thus we conclude that the case for motor inhibition affecting  
463 behavioural flexibility in the context of problem solving remains absent, but it remains possible  
464 that it reflects other facets of behavioural flexibility (reviewed in Bari and Robbins, 2013).

#### 465 *Previous experience*

466 Birds with previous experience of having solved the opaque lever device, or indeed any of the  
467 three access points during the main trials, were more accurate and had higher solving success  
468 in subsequent trials. Furthermore, performance improved with experience over repeated  
469 problem-solving attempts with regards to that particular solving method, perhaps owing to  
470 instrumental conditioning. Thus, attributing an individual's cognitive performance to how  
471 quickly they solve a problem, or their ability to solve multiple novel problems may be a  
472 function of their previous experience (Rowe & Healy, 2014; Sih & Del Giudice, 2012). While  
473 we acknowledge the constraints in controlling for all experiences animals may have had with  
474 features of an experimental apparatus, especially if based on simple generalizable rules, tasks  
475 could be designed such that they include multiple access points that vary in modality (e.g.  
476 sensory vs motor; smell and touch), that vary in the appearance of the specific materials they  
477 use (e.g. white plastic vs. black plastic), and/or motor skills as we have attempted to do here  
478 (Auersperg et al., 2011; Griffin & Guez, 2014; Manrique et al., 2013; Overington et al., 2009).  
479 This paradigm may facilitate the testing of true innovations that are not confounded by previous

480 experience, or alternatively, to explicitly test what kinds of experiences facilitate future  
481 innovations.

482 *Repeatability, pseudo-repeatability and positive feedback*

483 Our results demonstrate repeatable individual differences across two behaviours involved with  
484 problem-solving behaviour (latency to touch the device, and accuracy when interacting with  
485 the device), and for problem solving success itself. Adjusted and unadjusted repeatabilities  
486 differed for all three behaviours. For latency, repeatability decreased but remained significant  
487 after controlling for hunger-induced motivation, suggesting that some of the between-  
488 individual differences in the unadjusted repeatability were caused by hunger. In contrast, for  
489 accuracy, repeatability *increased* (and again remained significant) after controlling for the  
490 effects of previous experience, suggesting that some of the within-individual variation (the  
491 error component) in the unadjusted analysis was explained by previous experience. And for  
492 problem solving success, repeatability was lost after controlling for accuracy, hunger, and  
493 previous experience, i.e., consistent individual differences in problem solving performance  
494 were explained entirely by these three factors. Thus, repeatable problem-solving behaviour  
495 arose because of a complex set of interactions between different factors which themselves  
496 differed consistently between individuals.

497 The significance of these findings is tied to the nature of the specific factor involved. Firstly,  
498 in the case of hunger, designed to manipulate motivation, each individual only experienced one  
499 of two treatments, a potentially reversible effect, suggesting that the component of the  
500 unadjusted between individual difference explained by hunger was inflated resulting in  
501 pseudo-repeatability. Although some sources of motivation are likely permanent, either  
502 through permanent environment (Wilson, 2018) or intrinsic motivation (Ebel & Call, 2018;  
503 Gajdon et al., 2014; Polizzi di Sorrentino et al., 2014; Taffoni et al., 2014), this pseudo measure

504 demonstrates that failure to control for motivation caused by temporary factors can inflate the  
505 intrinsic between-individual differences that researchers are attempting to characterise, that is,  
506 those differences that are caused by permanent environment or intrinsic effects. Secondly,  
507 accuracy explained some of the between individual variation, suggesting that the mechanisms  
508 underlying accurate interaction with the device, vary consistently between individuals  
509 themselves, and explain some of the between individual differences in the problem-solving  
510 performance. It appears likely these mechanisms are intrinsic rather than reversible, since  
511 motivation is controlled for in these analyses. Thirdly, experience also caused some of the  
512 between individual differences in problem solving performance, and since experience is not  
513 reversible, and by definition, carries forward into the next stage of the sequential problem  
514 solving process, this suggests a positive feedback loop driving consistency between individual  
515 differences in problem solving behaviour. Although the role for feedback loops driving  
516 differences in individual behaviour is well known (Dall et al., 2004; Sih et al., 2015), and  
517 examples of positive feedbacks are common in nature (Kishida et al., 2011), to our knowledge  
518 none have explained consistent between-individual differences. In this case we assume the  
519 feedback caused by experience leads to a permanent effect, though it remains possible that  
520 individuals eventually forget the experience.

521 Our results highlight the challenges of characterising consistent individual variation in  
522 sequential problem-solving performance as a measure of overall innovativeness. More  
523 generally, they provide a demonstration of how between-individual differences in innovation  
524 can be explained by inflated estimates of within-individual variation in motivation, inflated  
525 between-individual variation in accuracy, and by feedback loops involving previous  
526 experience. Much of the focus in studies on the evolutionary ecology of behaviour in general  
527 has been on the evolutionary processes that drive intrinsic individual variation. Our data



528 supports the idea that complex sources of variation underlying single traits are likely to make  
529 predicting the selective consequences of this variation challenging.

#### 530 ACKNOWLEDGEMENTS

531 We would like to thank Anouk de Plaa, Ruben de Wit, Marylou Aaldering and Franca  
532 Kropman for taking care of the birds and helping with the experiments. We thank Rémi Pitiot  
533 for coding videos for our inter-observer reliability analysis, and Mark Whittingham for advice  
534 on the analysis and Anthony Caravaggi for his assistance in generating the graphs. We thank  
535 Lucy Aplin and four anonymous referees for their constructive comments and suggestions  
536 during the review process. Funding for ACC and GLD from the European Research Council  
537 under the European Union’s Horizon 2020 Programme (FP7/2007-2013)/ERC Consolidator  
538 Grant “EVOLECOCOG” Project No. 617509, awarded to JLQ, and by a Science Foundation  
539 Ireland ERC Support Grant 14/ERC/B3118 to JLQ.

#### 540 REFERENCES

- 541 Amy, M., van Oers, K., & Naguib, M. (2012). Worms under cover: Relationships between  
542 performance in learning tasks and personality in great tits (*Parus major*). *Animal*  
543 *Cognition*, *15*(5), 763–770. <https://doi.org/10.1007/s10071-012-0500-3>
- 544 Aplin, L. M., Farine, D. R., Mann, R. P., & Sheldon, B. C. (2014). Individual-level  
545 personality influences social foraging and collective behaviour in wild birds.  
546 *Proceedings of the Royal Society B: Biological Sciences*, *281*(1789), 20141016.  
547 <https://doi.org/10.1098/rspb.2014.1016>
- 548 Auersperg, A. M. I., Bayern, A. M. P. von, Gajdon, G. K., Huber, L., & Kacelnik, A. (2011).  
549 Flexibility in Problem Solving and Tool Use of Kea and New Caledonian Crows in a  
550 Multi Access Box Paradigm. *PLOS ONE*, *6*(6), e20231.  
551 <https://doi.org/10.1371/journal.pone.0020231>

552 Auersperg, A. M. I., Gajdon, G. K., & Bayern, A. M. P. von. (2012). A new approach to  
553 comparing problem solving, flexibility and innovation. *Communicative & Integrative*  
554 *Biology*, 5(2), 140–145. <https://doi.org/10.4161/cib.18787>

555 Bari, A., & Robbins, T. W. (2013). Inhibition and impulsivity: Behavioral and neural basis of  
556 response control. *Progress in Neurobiology*, 108, 44–79.  
557 <https://doi.org/10.1016/j.pneurobio.2013.06.005>

558 Barrett, L. P., Stanton, L. A., & Benson-Amram, S. (2019). The cognition of ‘nuisance’  
559 species. *Animal Behaviour*, 147, 167–177.  
560 <https://doi.org/10.1016/j.anbehav.2018.05.005>

561 Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects  
562 Models Using lme4. *Journal of Statistical Software*, 67(1), 1–48.  
563 <https://doi.org/10.18637/jss.v067.i01>

564 Benson-Amram, S., Dantzer, B., Stricker, G., Swanson, E. M., & Holekamp, K. E. (2016).  
565 Brain size predicts problem-solving ability in mammalian carnivores. *Proceedings of*  
566 *the National Academy of Sciences*, 113(9), 2532–2537.  
567 <https://doi.org/10.1073/pnas.1505913113>

568 Benson-Amram, S., & Holekamp, K. E. (2012). Innovative problem solving by wild spotted  
569 hyenas. *Proc. R. Soc. B*, 279(1744), 4087–4095.  
570 <https://doi.org/10.1098/rspb.2012.1450>

571 Beran, M. J. (2015). The comparative science of “self-control”: What are we talking about?  
572 *Frontiers in Psychology*, 6. <https://doi.org/10.3389/fpsyg.2015.00051>

573 Birch, H. G. (1945). The role of motivational factors in insightful problem-solving. *Journal*  
574 *of Comparative Psychology*, 38(5), 295–317. <https://doi.org/10.1037/h0059937>

575 Boogert, N. J., Anderson, R. C., Peters, S., Searcy, W. A., & Nowicki, S. (2011). Song  
576 repertoire size in male song sparrows correlates with detour reaching, but not with

577 other cognitive measures. *Animal Behaviour*, 81(6), 1209–1216.  
578 <https://doi.org/10.1016/j.anbehav.2011.03.004>

579 Catry, P., Ruxton, G. D., Ratcliffe, N., Hamer, K. C., & Furness, R. W. (1999). Short-Lived  
580 Repeatabilities in Long-Lived Great Skuas: Implications for the Study of Individual  
581 Quality. *Oikos*, 84(3), 473–479. JSTOR. <https://doi.org/10.2307/3546426>

582 Cauchard, L., Boogert, N. J., Lefebvre, L., Dubois, F., & Doligez, B. (2013). Problem-  
583 solving performance is correlated with reproductive success in a wild bird population.  
584 *Animal Behaviour*, 85(1), 19–26. <https://doi.org/10.1016/j.anbehav.2012.10.005>

585 Cauchoix M., Chow P. K. Y., van Horik J. O., Atance C. M., Barbeau E. J., Barragan-Jason  
586 G., Bize P., Boussard A., Buechel S. D., Cabirol A., Cauchard L., Claidière N.,  
587 Dalesman S., Devaud J. M., Didic M., Doligez B., Fagot J., Fichtel C., Henke-von der  
588 Malsburg J., ... Morand-Ferron J. (2018). The repeatability of cognitive performance:  
589 A meta-analysis. *Philosophical Transactions of the Royal Society B: Biological  
590 Sciences*, 373(1756), 20170281. <https://doi.org/10.1098/rstb.2017.0281>

591 Cole, E. F., Cram, D. L., & Quinn, J. L. (2011). Individual variation in spontaneous problem-  
592 solving performance among wild great tits. *Animal Behaviour*, 81(2), 491–498.  
593 <https://doi.org/10.1016/j.anbehav.2010.11.025>

594 Cole, E. F., Morand-Ferron, J., Hinks, A. E., & Quinn, J. L. (2012). Cognitive Ability  
595 Influences Reproductive Life History Variation in the Wild. *Current Biology*, 22(19),  
596 1808–1812. <https://doi.org/10.1016/j.cub.2012.07.051>

597 Cole, E. F., & Quinn, J. L. (2012). Personality and problem-solving performance explain  
598 competitive ability in the wild. *Proceedings of the Royal Society of London B:  
599 Biological Sciences*, rspb20111539. <https://doi.org/10.1098/rspb.2011.1539>

600 Cole, E. F., & Quinn, J. L. (2014). Shy birds play it safe: Personality in captivity predicts risk  
601 responsiveness during reproduction in the wild. *Biology Letters*, *10*(5), 20140178.  
602 <https://doi.org/10.1098/rsbl.2014.0178>

603 Coppens, C. M., de Boer Sietse F., & Koolhaas Jaap M. (2010). Coping styles and  
604 behavioural flexibility: Towards underlying mechanisms. *Philosophical Transactions*  
605 *of the Royal Society B: Biological Sciences*, *365*(1560), 4021–4028.  
606 <https://doi.org/10.1098/rstb.2010.0217>

607 Dall, S. R. X., Houston, A. I., & McNamara, J. M. (2004). The behavioural ecology of  
608 personality: Consistent individual differences from an adaptive perspective. *Ecology*  
609 *Letters*, *7*(8), 734–739. <https://doi.org/10.1111/j.1461-0248.2004.00618.x>

610 Daniels, S. E., Fanelli, R. E., Gilbert, A., & Benson-Amram, S. (2019). Behavioral flexibility  
611 of a generalist carnivore. *Animal Cognition*. [https://doi.org/10.1007/s10071-019-](https://doi.org/10.1007/s10071-019-01252-7)  
612 [01252-7](https://doi.org/10.1007/s10071-019-01252-7)

613 David, M., Auclair, Y., & Cézilly, F. (2012). Assessing Short- and Long-Term Repeatability  
614 and Stability of Personality in Captive Zebra Finches Using Longitudinal Data.  
615 *Ethology*, *118*(10), 932–942. <https://doi.org/10.1111/j.1439-0310.2012.02085.x>

616 Dingemanse, N. J., Both, C., Drent, P. J., van Oers, K., & van Noordwijk, A. J. (2002).  
617 Repeatability and heritability of exploratory behaviour in great tits from the wild.  
618 *Animal Behaviour*, *64*(6), 929–938. <https://doi.org/10.1006/anbe.2002.2006>

619 Dingemanse, N. J., & Dochtermann, N. A. (2013). Quantifying individual variation in  
620 behaviour: Mixed-effect modelling approaches. *Journal of Animal Ecology*, *82*(1),  
621 39–54. <https://doi.org/10.1111/1365-2656.12013>

622 Drent, P. J., Oers, K. van, & Noordwijk, A. J. van. (2003). Realized heritability of  
623 personalities in the great tit (*Parus major*). *Proceedings of the Royal Society of*

624           *London. Series B: Biological Sciences*, 270(1510), 45–51.  
625           <https://doi.org/10.1098/rspb.2002.2168>

626   Ducatez, S., Audet, J. N., & Lefebvre, L. (2015). Problem-solving and learning in Carib  
627           grackles: Individuals show a consistent speed–accuracy trade-off. *Animal Cognition*,  
628           18(2), 485–496. <https://doi.org/10.1007/s10071-014-0817-1>

629   Dunn, J. C., Cole, E. F., & Quinn, J. L. (2011). Personality and parasites: Sex-dependent  
630           associations between avian malaria infection and multiple behavioural traits.  
631           *Behavioral Ecology and Sociobiology*, 65(7), 1459–1471.  
632           <https://doi.org/10.1007/s00265-011-1156-8>

633   Ebel, S. J., & Call, J. (2018). The Interplay of Prior Experience and Motivation in Great Ape  
634           Problem-Solving (Gorilla gorilla, Pan paniscus, Pan troglodytes, and Pongo abelii).  
635           *Journal of Comparative Psychology*, 132(3), 294–305.

636   Fragaszy, D. M., Biro, D., Eshchar, Y., Humle, T., Izar, P., Resende, B., & Visalberghi, E.  
637           (2013). The fourth dimension of tool use: Temporally enduring artefacts aid primates  
638           learning to use tools. *Philosophical Transactions of the Royal Society B: Biological*  
639           *Sciences*, 368(1630), 20120410. <https://doi.org/10.1098/rstb.2012.0410>

640   Friard, O., & Gamba, M. (2016). BORIS: A free, versatile open-source event-logging  
641           software for video/audio coding and live observations. *Methods in Ecology and*  
642           *Evolution*, 7(11), 1325–1330. <https://doi.org/10.1111/2041-210X.12584>

643   Gajdon, G. K., Lichtnegger, M., & Huber, L. (2014). *What a Parrot’s Mind Adds to Play:*  
644           *The Urge to Produce Novelty Fosters Tool Use Acquisition in Kea.*

645   Greenberg, R. (2003). The Role of Neophobia and Neophilia in the Development of  
646           Innovative Behaviour of Birds. In Simon M. Reader & K. N. Laland (Eds.), *Animal*  
647           *Innovation* (pp. 175–196). Oxford University Press.  
648           <https://doi.org/10.1093/acprof:oso/9780198526223.003.0008>

- 649 Greenberg, R. S., & Mettke-Hofmann, C. (2001). *Ecological aspects of neophobia and*  
650 *neophilia in birds*. <http://repository.si.edu/handle/10088/8053>
- 651 Griffin, A. S., & Diquelou, M. C. (2015). Innovative problem solving in birds: A cross-  
652 species comparison of two highly successful passerines. *Animal Behaviour*, *100*, 84–  
653 94. <https://doi.org/10.1016/j.anbehav.2014.11.012>
- 654 Griffin, A. S., Diquelou, M., & Perea, M. (2014). Innovative problem solving in birds: A key  
655 role of motor diversity. *Animal Behaviour*, *92*, 221–227.  
656 <https://doi.org/10.1016/j.anbehav.2014.04.009>
- 657 Griffin, A. S., & Guez, D. (2014). Innovation and problem solving: A review of common  
658 mechanisms. *Behavioural Processes*, *109, Part B*, 121–134.  
659 <https://doi.org/10.1016/j.beproc.2014.08.027>
- 660 Hämäläinen, L., Rowland, H. M., Mappes, J., & Thorogood, R. (2019). The effect of social  
661 information from live demonstrators compared to video playback on blue tit foraging  
662 decisions. *PeerJ*, *7*, e7998. <https://doi.org/10.7717/peerj.7998>
- 663 Hartig, F. (2020). *DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed)*  
664 *Regression Models* (R package version 0.3.3.0.) [Computer software].  
665 <https://CRAN.R-project.org/package=DHARMa>
- 666 Highcock, L., & Carter, A. J. (2014). Intraindividual Variability of Boldness Is Repeatable  
667 across Contexts in a Wild Lizard. *PLOS ONE*, *9*(4), e95179.  
668 <https://doi.org/10.1371/journal.pone.0095179>
- 669 Hopper, L. M., Price, S. A., Freeman, H. D., Lambeth, S. P., Schapiro, S. J., & Kendal, R. L.  
670 (2014). Influence of personality, age, sex, and estrous state on chimpanzee problem-  
671 solving success. *Animal Cognition*, *17*(4), 835–847. [https://doi.org/10.1007/s10071-](https://doi.org/10.1007/s10071-013-0715-y)  
672 [013-0715-y](https://doi.org/10.1007/s10071-013-0715-y)

673 Johnson-Ulrich, L., Johnson-Ulrich, Z., & Holekamp, K. (2018). Proactive behavior, but not  
674 inhibitory control, predicts repeated innovation by spotted hyenas tested with a multi-  
675 access box. *Animal Cognition*, *21*(3), 379–392. [https://doi.org/10.1007/s10071-018-](https://doi.org/10.1007/s10071-018-1174-2)  
676 [1174-2](https://doi.org/10.1007/s10071-018-1174-2)

677 Jolles, J. W., Briggs, H. D., Araya-Ajoy, Y. G., & Boogert, N. J. (2019). Personality,  
678 plasticity and predictability in sticklebacks: Bold fish are less plastic and more  
679 predictable than shy fish. *Animal Behaviour*, *154*, 193–202.  
680 <https://doi.org/10.1016/j.anbehav.2019.06.022>

681 Kabadayi, C., Bobrowicz, K., & Osvath, M. (2018). The detour paradigm in animal  
682 cognition. *Animal Cognition*, *21*(1), 21–35. [https://doi.org/10.1007/s10071-017-1152-](https://doi.org/10.1007/s10071-017-1152-0)  
683 [0](https://doi.org/10.1007/s10071-017-1152-0)

684 Kishida, O., Trussell, G. C., Ohno, A., Kuwano, S., Ikawa, T., & Nishimura, K. (2011).  
685 Predation risk suppresses the positive feedback between size structure and  
686 cannibalism. *Journal of Animal Ecology*, *80*(6), 1278–1287.  
687 <https://doi.org/10.1111/j.1365-2656.2011.01871.x>

688 Kotrschal, A., & Taborsky, B. (2010). Environmental Change Enhances Cognitive Abilities  
689 in Fish. *PLOS Biology*, *8*(4), e1000351. <https://doi.org/10.1371/journal.pbio.1000351>

690 Laland, K. N., & Reader, S. M. (1999). Foraging innovation in the guppy. *Animal Behaviour*,  
691 *57*(2), 331–340. <https://doi.org/10.1006/anbe.1998.0967>

692 Lefebvre, L., Reader, S. M., & Sol, D. (2004). Brains, Innovations and Evolution in Birds and  
693 Primates. *Brain, Behavior and Evolution*, *63*(4), 233–246.  
694 <https://doi.org/10.1159/000076784>

695 Lefebvre, L., Whittle, P., Lascaris, E., & Finkelstein, A. (1997). Feeding innovations and  
696 forebrain size in birds. *Animal Behaviour*, *53*(3), 549–560.  
697 <https://doi.org/10.1006/anbe.1996.0330>

698 Lermite, F., Peneaux, C., & Griffin, A. S. (2017). Personality and problem-solving in  
699 common mynas (*Acridotheres tristis*). *Behavioural Processes*, *134*, 87–94.  
700 <https://doi.org/10.1016/j.beproc.2016.09.013>

701 Logan, C. J. (2016a). How far will a behaviourally flexible invasive bird go to innovate?  
702 *Royal Society Open Science*, *3*(6), 160247. <https://doi.org/10.1098/rsos.160247>

703 Logan, C. J. (2016b). Behavioral flexibility and problem solving in an invasive bird. *PeerJ*, *4*,  
704 e1975. <https://doi.org/10.7717/peerj.1975>

705 Logan, C. J. (2016c). Behavioral flexibility in an invasive bird is independent of other  
706 behaviors. *PeerJ*, *4*, e2215. <https://doi.org/10.7717/peerj.2215>

707 MacLean, E. L., Hare, B., Nunn, C. L., Addessi, E., Amici, F., Anderson, R. C., Aureli, F.,  
708 Baker, J. M., Bania, A. E., Barnard, A. M., Boogert, N. J., Brannon, E. M., Bray, E.  
709 E., Bray, J., Brent, L. J. N., Burkart, J. M., Call, J., Cantlon, J. F., Cheke, L. G., ...  
710 Zhao, Y. (2014). The evolution of self-control. *Proceedings of the National Academy*  
711 *of Sciences*, *111*(20), E2140–E2148. <https://doi.org/10.1073/pnas.1323533111>

712 Manrique, H. M., Völter, C. J., & Call, J. (2013). Repeated innovation in great apes. *Animal*  
713 *Behaviour*, *85*(1), 195–202. <https://doi.org/10.1016/j.anbehav.2012.10.026>

714 Mateos-Gonzalez, F., Quesada, J., & Senar, J. C. (2011). Sexy birds are superior at solving a  
715 foraging problem. *Biology Letters*, *7*(5), 668–669.  
716 <https://doi.org/10.1098/rsbl.2011.0163>

717 Morand- Ferron, J., Cole, E. F., & Quinn, J. L. (2016). Studying the evolutionary ecology of  
718 cognition in the wild: A review of practical and conceptual challenges. *Biological*  
719 *Reviews*, *91*(2), 367–389. <https://doi.org/10.1111/brv.12174>

720 Morand-Ferron, J., Cole, E. F., Rawles, J. E. C., & Quinn, J. L. (2011). Who are the  
721 innovators? A field experiment with 2 passerine species. *Behavioral Ecology*, *22*(6),  
722 1241–1248. <https://doi.org/10.1093/beheco/arr120>



723 Morand-Ferron, J., & Quinn, J. L. (2011). Larger groups of passerines are more efficient  
724 problem solvers in the wild. *Proceedings of the National Academy of Sciences*,  
725 *108*(38), 15898–15903. <https://doi.org/10.1073/pnas.1111560108>

726 Morton, F. B., Lee, P. C., & Buchanan-Smith, H. M. (2013). Taking personality selection  
727 bias seriously in animal cognition research: A case study in capuchin monkeys  
728 (*Sapajus apella*). *Animal Cognition*, *16*(4), 677–684. [https://doi.org/10.1007/s10071-](https://doi.org/10.1007/s10071-013-0603-5)  
729 [013-0603-5](https://doi.org/10.1007/s10071-013-0603-5)

730 Overington, S. E., Cauchard, L., Côté, K.-A., & Lefebvre, L. (2011). Innovative foraging  
731 behaviour in birds: What characterizes an innovator? *Behavioural Processes*, *87*(3),  
732 274–285. <https://doi.org/10.1016/j.beproc.2011.06.002>

733 Overington, S. E., Morand-Ferron, J., Boogert, N. J., & Lefebvre, L. (2009). Technical  
734 innovations drive the relationship between innovativeness and residual brain size in  
735 birds. *Animal Behaviour*, *78*(4), 1001–1010.  
736 <https://doi.org/10.1016/j.anbehav.2009.06.033>

737 Pecora, G., Zoratto, F., Paoletti, M., Bellagamba, F., Paglieri, F., & Addessi, E. (2017).  
738 Executive Function. In J. Vonk & T. Shackelford (Eds.), *Encyclopedia of Animal*  
739 *Cognition and Behavior* (pp. 1–5). Springer International Publishing.  
740 [https://doi.org/10.1007/978-3-319-47829-6\\_1048-1](https://doi.org/10.1007/978-3-319-47829-6_1048-1)

741 Pinheiro, J, Bates, D., DebRoy, S., Sarkar, D., & R Core Team. (2019). *nlme: Linear and*  
742 *Nonlinear Mixed Effects Models* (3.1-141) [Computer software]. [https://CRAN.R-](https://CRAN.R-project.org/package=nlme)  
743 [project.org/package=nlme](https://CRAN.R-project.org/package=nlme)

744 Pinheiro, José, & Bates, D. (2006). *Mixed-Effects Models in S and S-PLUS*. Springer Science  
745 & Business Media.

746 Polizzi di Sorrentino, E., Sabbatini, G., Truppa, V., Bordonali, A., Taffoni, F., Formica, D.,  
747 Baldassarre, G., Mirolli, M., Guglielmelli, E., & Visalberghi, E. (2014). Exploration

748 and learning in capuchin monkeys (*Sapajus* spp.): The role of action–outcome  
749 contingencies. *Animal Cognition*, 17(5), 1081–1088. [https://doi.org/10.1007/s10071-](https://doi.org/10.1007/s10071-014-0740-5)  
750 014-0740-5

751 Preiszner, B., Papp, S., Pipoly, I., Seress, G., Vincze, E., Liker, A., & Bókony, V. (2017).  
752 Problem-solving performance and reproductive success of great tits in urban and  
753 forest habitats. *Animal Cognition*, 20(1), 53–63. [https://doi.org/10.1007/s10071-016-](https://doi.org/10.1007/s10071-016-1008-z)  
754 1008-z

755 Reader, S. (2003). Innovation and social learning: Individual variation and brain evolution.  
756 *Animal Biology*, 53(2), 147–158. <https://doi.org/10.1163/157075603769700340>

757 Reader, Simon M., & Laland, K. N. (2002). Social intelligence, innovation, and enhanced  
758 brain size in primates. *Proceedings of the National Academy of Sciences*, 99(7),  
759 4436–4441. <https://doi.org/10.1073/pnas.062041299>

760 Reader, S.M., & Laland, K. N. (2003). *Animal Innovation*. Oxford University Press.

761 Réale, D., Reader, S. M., Sol, D., McDougall, P. T., & Dingemanse, N. J. (2007). Integrating  
762 animal temperament within ecology and evolution. *Biological Reviews*, 82(2), 291–  
763 318. <https://doi.org/10.1111/j.1469-185X.2007.00010.x>

764 Rothbart, M. K., Posner, M. I., & Telzrow, C. F. (1985). Temperament and the development  
765 of selfregulation. In *The neuropsychology of individual differences: A developmental*  
766 *perspective*. (pp. 93–125). Plenum.

767 Rowe, C., & Healy, S. D. (2014). Measuring variation in cognition. *Behavioral Ecology*,  
768 25(6), 1287–1292. <https://doi.org/10.1093/beheco/aru090>

769 RStudio Team. (2019). *R: A language and environment for statistical computing*. R  
770 *Foundation for Statistical Computing*. RStudio, Inc. <https://www.rstudio.org/>

771 Seed, A. M., & Call, J. (2010). Problem-solving in tool-using and non-tool-using animals. In  
772 *Encyclopedia of animal behavior* (pp. 778–785).

773 Seed, A. M., & Mayer, C. M. P. (2017). Problem Solving. In *APA Handbook of Comparative*  
774 *Psychology: Vol. 2. Perception, Learning, and Cognition: APA*. American  
775 Psychological Association.

776 Sih, A., & Del Giudice, M. (2012). Linking behavioural syndromes and cognition: A  
777 behavioural ecology perspective. *Philosophical Transactions of the Royal Society B:*  
778 *Biological Sciences*, 367(1603), 2762–2772. <https://doi.org/10.1098/rstb.2012.0216>

779 Sih, A., Mathot, K. J., Moirón, M., Montiglio, P.-O., Wolf, M., & Dingemanse, N. J. (2015).  
780 Animal personality and state-behaviour feedbacks: A review and guide for  
781 empiricists. *Trends in Ecology & Evolution*, 30(1), 50–60.  
782 <https://doi.org/10.1016/j.tree.2014.11.004>

783 Smit, J. A. H., & van Oers, K. (2019). Personality types vary in their personal and social  
784 information use. *Animal Behaviour*. <https://doi.org/10.1016/j.anbehav.2019.02.002>

785 Sol, D., Duncan, R. P., Blackburn, T. M., Cassey, P., & Lefebvre, L. (2005). Big brains,  
786 enhanced cognition, and response of birds to novel environments. *Proceedings of the*  
787 *National Academy of Sciences*, 102(15), 5460–5465.  
788 <https://doi.org/10.1073/pnas.0408145102>

789 Sol, D., Griffin, A. S., & Bartomeus, I. (2012). Consumer and motor innovation in the  
790 common myna: The role of motivation and emotional responses. *Animal Behaviour*,  
791 83(1), 179–188. <https://doi.org/10.1016/j.anbehav.2011.10.024>

792 Sol, D., Griffin, A. S., Bartomeus, I., & Boyce, H. (2011). Exploring or Avoiding Novel Food  
793 Resources? The Novelty Conflict in an Invasive Bird. *PLOS ONE*, 6(5), e19535.  
794 <https://doi.org/10.1371/journal.pone.0019535>

795 Sol, D., Lefebvre, L., & Rodríguez-Teijeiro, J. D. (2005). Brain size, innovative propensity  
796 and migratory behaviour in temperate Palaearctic birds. *Proceedings of the Royal*

797            *Society B: Biological Sciences*, 272(1571), 1433–1441.  
798            <https://doi.org/10.1098/rspb.2005.3099>

799 Stoffel, M. A., Nakagawa, S., & Schielzeth, H. (2017). rptR: Repeatability estimation and  
800            variance decomposition by generalized linear mixed-effects models. *Methods in*  
801            *Ecology and Evolution*, 8(11), 1639–1644. <https://doi.org/10.1111/2041-210X.12797>

802 Suomi, S. J. (2004). How Gene-Environment Interactions Shape Biobehavioral Development:  
803            Lessons From Studies With Rhesus Monkeys. *Research in Human Development*, 1(3),  
804            205–222. [https://doi.org/10.1207/s15427617rhd0103\\_5](https://doi.org/10.1207/s15427617rhd0103_5)

805 Taffoni, F., Tamilia, E., Focaroli, V., Formica, D., Ricci, L., Di Pino, G., Baldassarre, G.,  
806            Mirolli, M., Guglielmelli, E., & Keller, F. (2014). Development of goal-directed  
807            action selection guided by intrinsic motivations: An experiment with children.  
808            *Experimental Brain Research*, 232(7), 2167–2177. [https://doi.org/10.1007/s00221-](https://doi.org/10.1007/s00221-014-3907-z)  
809            014-3907-z

810 Taylor, A. H., Hunt, G. R., Medina, F. S., & Gray, R. D. (2009). Do New Caledonian crows  
811            solve physical problems through causal reasoning? *Proceedings of the Royal Society*  
812            *B: Biological Sciences*, 276(1655), 247–254. <https://doi.org/10.1098/rspb.2008.1107>

813 Thorndike, E. (1911). *Animal intelligence: Experimental studies*. The Macmillan Company.

814 Thornton, A., & Samson, J. (2012). Innovative problem solving in wild meerkats. *Animal*  
815            *Behaviour*, 83(6), 1459–1468. <https://doi.org/10.1016/j.anbehav.2012.03.018>

816 Trompf, L., & Brown, C. (2014). Personality affects learning and trade-offs between private  
817            and social information in guppies, *Poecilia reticulata*. *Animal Behaviour*, 88, 99–106.  
818            <https://doi.org/10.1016/j.anbehav.2013.11.022>

819 van Horik, J. O., Langley, E. J. G., Whiteside, M. A., Laker, P. R., Beardsworth, C. E., &  
820            Madden, J. R. (2018). Do detour tasks provide accurate assays of inhibitory control?

821 *Proceedings of the Royal Society B: Biological Sciences*, 285(1875), 20180150.  
822 <https://doi.org/10.1098/rspb.2018.0150>

823 van Horik, J. O., & Madden, J. R. (2016). A problem with problem solving: Motivational  
824 traits, but not cognition, predict success on novel operant foraging tasks. *Animal*  
825 *Behaviour*, 114, 189–198. <https://doi.org/10.1016/j.anbehav.2016.02.006>

826 van Oers, K., de Jong, G., van Noordwijk, A. J., Kempenaers, B., & Drent, P. J. (2005).  
827 Contribution of Genetics to the Study of Animal Personalities: A Review of Case  
828 Studies. *Behaviour*, 142(9/10), 1185–1206.

829 van Oers, K., Drent, P. J., de Goede, P., & van Noordwijk, A. J. (2004). Realized heritability  
830 and repeatability of risk-taking behaviour in relation to avian personalities.  
831 *Proceedings of the Royal Society of London. Series B: Biological Sciences*,  
832 271(1534), 65–73. <https://doi.org/10.1098/rspb.2003.2518>

833 van Oers, K., & Naguib, M. (2013). Avian Personality. In *Animal Personalities: Behavior,*  
834 *Physiology, and Evolution. C. Carere and D. Maestripieri* (pp. 66–95). The  
835 University Chicago Press.

836 Verbeek, M. E. M., Drent, P. J., & Wiepkema, P. R. (1994). Consistent individual differences  
837 in early exploratory behaviour of male great tits. *Animal Behaviour*, 48(5), 1113–  
838 1121. <https://doi.org/10.1006/anbe.1994.1344>

839 von Bayern, A. M. P., Heathcote, R. J. P., Rutz, C., & Kacelnik, A. (2009). The Role of  
840 Experience in Problem Solving and Innovative Tool Use in Crows. *Current Biology*,  
841 19(22), 1965–1968. <https://doi.org/10.1016/j.cub.2009.10.037>

842 Webster, S. J., & Lefebvre, L. (2001). Problem solving and neophobia in a columbiform–  
843 passeriform assemblage in Barbados. *Animal Behaviour*, 62(1), 23–32.  
844 <https://doi.org/10.1006/anbe.2000.1725>

845 Westneat, D. F., Hatch, M. I., Wetzel, D. P., & Ensminger, A. L. (2011). Individual variation  
846 in parental care reaction norms: Integration of personality and plasticity. *The*  
847 *American Naturalist*, 178(5), 652–667. <https://doi.org/10.1086/662173>

848 Whittingham, M. J., Stephens, P. A., Bradbury, R. B., & Freckleton, R. P. (2006). Why do  
849 we still use stepwise modelling in ecology and behaviour? *Journal of Animal*  
850 *Ecology*, 75(5), 1182–1189. <https://doi.org/10.1111/j.1365-2656.2006.01141.x>

851 Wilson, A. J. (2018). How should we interpret estimates of individual repeatability?  
852 *Evolution Letters*, 2(1), 4–8. <https://doi.org/10.1002/evl3.40>

853 Zandberg, L., Quinn, J. L., Naguib, M., & van Oers, K. (2017). Personality-dependent  
854 differences in problem-solving performance in a social context reflect foraging  
855 strategies. *Behavioural Processes*, 134, 95–102.  
856 <https://doi.org/10.1016/j.beproc.2016.09.007>

857

858 Table 1 Full model outputs from GLMM with factors affecting latency to touch the device per  
859 trial ( $N = 226$ , d.f. = 190,  $R^2$  (marginal = 0.26, conditional = 0.49, AIC = 739.83). Random  
860 effect included bird identity (0.72; 95%CI (0.51, 1.02)) and the residual variance (1.08;  
861 95% C.I. (0.97, 1.19)).

<b>Fixed effects</b>	<b><math>\beta \pm \text{s.e.}</math></b>	<b>t</b>	<b>P</b>
Intercept	4.27 $\pm$ 0.35	12.19	<0.001
<b>Motivation group<sup>a</sup></b>	<b>1.17 <math>\pm</math> 0.30</b>	<b>3.84</b>	<b>&lt;0.001</b>
Personality <sup>b</sup>	0.20 $\pm$ 0.30	0.67	0.507
<b>Trial number</b>	<b>-0.09 <math>\pm</math> 0.03</b>	<b>-3.36</b>	<b>0.001</b>

Sex<sup>c</sup> -0.53 ± 0.30 -1.76 0.089

Previous experience<sup>d</sup> -0.19 ± 0.17 -1.15 0.25

862 <sup>a</sup> low (reference level is high), <sup>b</sup> slow (reference level is fast), <sup>c</sup> male (reference level is female),  
863 <sup>d</sup> yes (reference level is no). Significant result ( $P < 0.05$ ) is highlighted in bold.

864 Table 2 Full model outputs from GLMM with factors affecting accuracy per trial ( $N = 222$ ,  
865 d.f.= 182,  $R^2$  [marginal = 0.33, conditional = 0.56, AIC = 209.4). Random effect included bird  
866 identity (0.24; 95%CI (0.22, 0.27)) and the residual variance (0.15; 95%CI (0.10, 0.23)).

Fixed	$\beta \pm \text{s.e.}$	t	P
Intercept	0.29 ± 0.08	3.44	<0.001
Interaction rate	0.00 ± 0.00	0.43	0.670
<b>Previous experience<sup>a</sup></b>	<b>0.35 ± 0.04</b>	<b>8.20</b>	<b>&lt;0.001</b>
Personality <sup>b</sup>	0.13 ± 0.07	1.80	0.082
Motivation group <sup>c</sup>	-0.04 ± 0.07	-0.54	0.591
Fused trial	-0.11 ± 0.07	-1.70	0.091
Sex	0.05 ± 0.07	0.64	0.524
Trial number	-0.00 ± 0.00	-0.33	0.744

867 <sup>a</sup> yes (reference level is no), <sup>b</sup> slow (reference level is fast), <sup>c</sup> low (reference level high).  
868 Significant result ( $P < 0.05$ ) is highlighted in bold.

869 Table 3 Full model outputs from GLMM with factors affecting solving within trials ( $N = 224$ ,  
 870  $d.f. = 216$ ,  $R^2$  [marginal = 0.51, conditional = 0.52, AIC = 55.85]. Random effect included bird  
 871 identity (0.37; 95%CI (0.41, 0.33)) and the residual variance (0.13; 95%CI (0.10, 0.27)).

872

Fixed effects	$\beta \pm s.e.$	$z$	$P$
Intercept	-1.93 $\pm$ 0.68	2.86	0.004
<b>Accuracy</b>	<b>3.79 <math>\pm</math> 0.78</b>	<b>4.90</b>	<b>&lt;0.001</b>
<b>Previous experience<sup>a</sup></b>	<b>1.26 <math>\pm</math> 0.47</b>	<b>2.66</b>	<b>0.008</b>
<b>Motivation group<sup>b</sup></b>	<b>-1.79 <math>\pm</math> 0.50</b>	<b>-3.57</b>	<b>&lt;0.001</b>
Personality <sup>c</sup>	-0.68 $\pm$ 0.43	-1.58	0.115
Sex <sup>d</sup>	0.26 $\pm$ 0.43	0.61	0.541
Trial number	0.09 $\pm$ 0.10	0.85	0.395
Fused trial	-0.18 $\pm$ 0.73	-0.25	0.806
Total number of touches per trial	-0.00 $\pm$ 0.00	-1.47	0.141

873 <sup>a</sup> yes (reference level is no), <sup>b</sup> low (reference level is high), <sup>c</sup> slow (reference level is fast), <sup>d</sup>  
 874 male (reference level is female), <sup>e</sup> yes (reference level is no). Significant result ( $P < 0.05$ ) is  
 875 highlighted in bold.

876 Table 4 Full model outputs from GLM with factors affecting the number of different access  
 877 points solved by an individual ( $N = 35$ ,  $d.f. = 24$ , pseudo- $R^2 = 0.24$ , AIC = 99.60).



<b>Explanatory variables</b>	<b><math>\beta \pm \text{s.e.}</math></b>	<b><i>z</i></b>	<b><i>P</i></b>
Intercept	<0.001 $\pm$ <0.001	1.84	0.066
<b>Motivation group<sup>a</sup></b>	<b>-1.04 <math>\pm</math> 0.37</b>	<b>-2.80</b>	<b>0.005</b>
Personality <sup>b</sup>	-0.01 $\pm$ 0.032	-0.30	0.761
Sex <sup>c</sup>	0.18 $\pm$ 0.32	0.57	0.567
Latency to touch the device (in first trial only)	<0.001 $\pm$ <0.001	-0.35	0.727

878 <sup>a</sup> low (reference level is high), <sup>b</sup> slow (reference level is fast), <sup>c</sup> male (reference level is female).

879 Significant result ( $P < 0.05$ ) is highlighted in bold.

880 Table 5. Repeatability (adjusted and unadjusted) estimates for the three main components of  
881 problem solving behaviour during the experiment. Unadjusted values are from mixed models  
882 with only individual as a random effect. Adjusted values also include significant fixed effects  
883 for each of a), b) and c) as shown in Tables 1, 2 and 3 respectively. In addition, for c), adjusted  
884 repeatabilities are also shown when single fixed effects were removed.

<b>model</b>	<b>repeatability</b>	<b><i>r</i></b>	<b>confidence interval</b>	<b><i>P</i></b>
		<b>value</b>	<b>(lower, upper)</b>	<b>value</b>
latency to touch the device	<b>adjusted</b>	<b>0.32</b>	<b>0.15, 0.45</b>	<b>&lt;0.001</b>
	<b>unadjusted</b>	<b>0.48</b>	<b>0.30, 0.61</b>	<b>&lt;0.001</b>
accuracy	<b>adjusted</b>	<b>0.45</b>	<b>0.17, 0.52</b>	<b>&lt;0.001</b>

	<b>unadjusted</b>	<b>0.23</b>	<b>0.08, 0.38</b>	<b>0.001</b>
solving within a trial	adjusted	0.03	0, 0.17	0.40
	<b>unadjusted</b>	<b>0.31</b>	<b>0.1, 0.46</b>	<b>&lt;0.001</b>
	<b>adjusted-without accuracy</b>	<b>0.18</b>	<b>0, 0.33</b>	<b>0.044</b>
	<b>adjusted-without motivation group</b>	<b>0.21</b>	<b>0, 0.35</b>	<b>0.017</b>
	adjusted-without previous experience	0.14	0, 0.29	0.076

885 Significant result ( $P < 0.05$ ) is highlighted in bold.

886

887

888 Table A1 Variable names and definitions

Variable	Definition
latency to touch	latency from start of trial to touch the device per trial.
personality	second and third generation of bi-directionally selected personality lines for fast and slow exploring personality
motivation	two motivation treatment groups, low motivation (sated birds fed 3 wax moth larvae 30mins before trial commences) or high motivation (birds that were food deprived 1hr before trial)
trial number	number of the trial

fused trial	when any access point fused was in the trial
bird identity	identity code for each bird
number of access points solved	absolute number of different access points on the device by individual across all trials
sex	male or female
inhibition score	number of trials until individual reaches criterion in detour reach task. (criterion = 4/5 consecutive successful trials)
solving within a trial	whether an individual has solved in a trial
accuracy	proportion of touches to functional access points on the device compared to all touches to the device per trial
total number of touches per trial	total number of touches to the device per trial
previous experience	whether in the current trial there was a functional, non-fused access point available to solve that they had solved previously; for the lever in the multi-access task, this included whether they had experience solving the opaque device
interaction rate	the number of total touches to any part of the device divided by duration interacting with the device (time from first touch to last touch per trial)

889

890 Table A2 Minimum adequate model outputs from GLMM with factors affecting latency to  
891 touch the device per trial ( $N = 226$ , d.f. = 190,  $R^2$  (marginal = 0.26, conditional = 0.47, AIC =  
892 726.50). Random effect included bird identity (1.24; 95%CI (1.08, 1.41)) and the residual  
893 variance (0.51; 95%CI (0.28, 0.95)).

894

Fixed effects	$\beta \pm \text{s.e.}$	t	P	variance
Intercept	4.25 $\pm$ 0.27	15.93	<0.001	
<b>Motivation group<sup>a</sup></b>	<b>1.19 <math>\pm</math> 0.29</b>	<b>4.10</b>	<b>&lt;0.001</b>	
<b>Trial number</b>	<b>-0.09 <math>\pm</math> 0.03</b>	<b>-3.30</b>	<b>0.001</b>	
Sex <sup>b</sup>	-0.55 $\pm$ 0.30	-1.95	0.06	

895 <sup>a</sup> low (reference level is high), <sup>b</sup> male (reference level is female). Significant result ( $P < 0.05$ )  
896 is highlighted in bold.

897

898 Table A3 Minimum adequate model outputs from GLMM with factors affecting accuracy per  
899 trial ( $N = 222$ , d.f. = 187,  $R^2$  [marginal = 0.34, conditional = 0.54, AIC = 49.42). Random  
900 effect included bird identity (0.24; 95% CI (0.20, 0.27)) and the residual variance (0.19;  
901 95% CI (0.13, 0.30).

Fixed effects	$\beta \pm \text{s.e.}$	t	P	variance
Intercept	0.27 $\pm$ 0.05	5.05	<0.001	
<b>Previous experience<sup>a</sup></b>	<b>0.36 <math>\pm</math> 0.04</b>	<b>8.80</b>	<b>&lt;0.001</b>	
Personality <sup>b</sup>	0.12 $\pm$ 0.06	1.88	0.068	
<b>Fused trial</b>	<b>-0.11 <math>\pm</math> 0.04</b>	<b>-2.55</b>	<b>0.012</b>	

902 <sup>a</sup> yes (reference level is no), <sup>b</sup> slow (reference level is fast). Significant result ( $P < 0.05$ ) is  
903 highlighted in bold.

904

905 Table A4 Minimum adequate model output from GLMM with factors affecting solving within  
906 trials ( $N = 224$ , d.f. = 223,  $R^2$  [marginal = 0.55, conditional = 0.58, AIC = 205.70). Random

907 effect included bird identity (0.37; 95% CI (0.33, 0.41)) and the residual variance (0.13; 95% CI  
 908 (0.09, 0.25)).

Fixed effects	$\beta \pm \text{s.e.}$	<i>z</i>	<i>P</i>	variance
Intercept	-2.42 ± 0.50	-4.88	<0.001	
<b>Accuracy</b>	<b>3.91 ± 0.78</b>	<b>4.99</b>	<b>&lt;0.001</b>	
<b>Previous experience<sup>a</sup></b>	<b>1.52 ± 0.42</b>	<b>3.65</b>	<b>&lt;0.001</b>	
<b>Motivation group<sup>b</sup></b>	<b>-1.64 ± 0.48</b>	<b>-3.36</b>	<b>&lt;0.001</b>	

909 <sup>a</sup> yes (reference level is no), <sup>b</sup> low (reference level is high). Significant result ( $P < 0.05$ ) is  
 910 highlighted in bold.

911

912 Table A5 Minimum adequate model outputs from GLM with factors affecting the number of  
 913 different access points solved by an individual ( $N = 35$ , d.f. = 33, pseudo- $R^2 = 0.23$ , AIC =  
 914 93.54).

Explanatory variables	$\beta \pm \text{s.e.}$	<i>z</i>	<i>P</i>
Intercept	0.54 0.18	3.03	<0.005
<b>Motivation group<sup>a</sup></b>	<b>-1.07 ± 0.36</b>	<b>-2.95</b>	<b>&lt;0.005</b>

919 <sup>a</sup> low (reference level is high). Significant result ( $P < 0.05$ ) is highlighted in bold.

920

921 Table A6 Full model outputs from GLMM with factors affecting latency to touch ( $N = 128$ ,  
 922 d.f.= 104). Data used in this model is from the subset of 22 individuals tested for inhibition  
 923 score. Random effect included bird identity (1.24, 95%CI (1.09, 1.42) and the residual variance  
 924 (0.51; 95%CI (0.28, 0.95)).

Fixed effects	$\beta \pm \text{s.e.}$	t	P	variance	
Intercept	4.69 $\pm$ 0.90	5.22	<0.001		925
<b>Motivation group<sup>a</sup></b>	<b>1.13 <math>\pm</math> 0.44</b>	<b>2.58</b>	<b>&lt;0.05</b>		926
Personality <sup>b</sup>	0.24 $\pm$ 0.46	0.51	0.62		927
Sex <sup>c</sup>	-0.49 $\pm$ 0.50	-0.97	0.35		928
<b>Trial number</b>	<b>-0.11 <math>\pm</math> 0.51</b>	<b>-2.09</b>	<b>&lt;0.05</b>		929
Previous experience <sup>d</sup>	-0.45 $\pm$ 0.28	-1.59	0.11		930
Inhibition score	-0.01 $\pm$ 0.07	-1.20	0.84		931
					932

933 <sup>a</sup> low (reference level is high), <sup>b</sup> slow (reference level is fast), <sup>c</sup> male (reference level is female).  
 934 Table A7 Full model outputs from GLMM with factors affecting accuracy per trial ( $N = 126$ ,  
 935 d.f.= 99). Data used in this model is from the subset of 22 individuals tested for inhibition  
 936 score. Random effect included bird identity (0.24; 95%CI (0.15, 0.36)) and the residual  
 937 variance (0.19; 95%CI (0.13, 0.30)).

938

Fixed effects	$\beta \pm \text{s.e.}$	t	P	variance
---------------	-------------------------	---	---	----------

Intercept	0.21 ± 0.23	0.88	0.38
Interaction rate	0.00 ± 0.00	0.81	0.42
<b>Previous experience</b>	<b>0.34 ± 0.07</b>	<b>4.67</b>	<b>&lt;0.001</b>
Personality <sup>a</sup>	0.16 ± 0.12	1.35	0.19
Motivation group <sup>b</sup>	-0.07 ± 0.11	-0.61	0.55
Fused trial	-0.19 ± 0.10	-1.89	0.06
Sex <sup>c</sup>	0.06 ± 0.13	0.49	0.62
Trial number	-0.00 ± 0.01	-0.29	0.76
Inhibition score	0.00 ± 0.02	0.31	0.75

939 <sup>a</sup> slow (reference level is fast), <sup>b</sup> low (reference level high), <sup>c</sup> male (reference level female).  
940 Table A8 Full model outputs from GLMM with factors affecting solving within trials ( $N = 126$ ,  
941 d.f. = 115). Data used in this model is from the subset of 22 individuals tested for inhibition  
942 score. Random effect included bird identity ( $<0.01$ ; 95%CI (0.0, 0.0)) and the residual variance  
943 ( $<0.01$ ; 95%CI (0.0, 0.0)).

<b>Fixed effects</b>	<b><math>\beta \pm \text{s.e.}</math></b>	<b><math>z</math></b>	<b><math>P</math></b>	<b>variance</b>
Intercept	-4.45 ± 1.49	-2.98	<0.001	
<b>Accuracy</b>	<b>4.36 ± 1.07</b>	<b>4.05</b>	<b>&lt;0.001</b>	
Previous experience <sup>a</sup>	1.38 ± 0.80	1.72	0.08	

<b>Motivation group<sup>b</sup></b>	<b>-2.14 ± 0.69</b>	<b>-3.10</b>	<b>&lt;0.01</b>
Personality <sup>c</sup>	-0.08 ± 0.64	0.13	0.90
Sex <sup>d</sup>	1.01 ± 0.73	1.38	0.17
Trial number	0.24 ± 0.16	1.44	0.15
Fused trial	-1.20 ± 1.22	-0.98	0.33
Total number of touches per trial	-0.00 ± 0.00	-1.37	0.17
Inhibition score	0.11 ± 0.10	1.11	0.28

944 <sup>a</sup> yes (reference level is no), <sup>b</sup> low (reference level is high), <sup>c</sup> slow (reference level is fast), <sup>d</sup>  
945 male (reference level is female).

946 Table A9 Full model outputs from GLM with factors affecting the number of different access  
947 points solved by an individual ( $N = 22$ , d.f. = 16).

<b>Explanatory variables</b>	<b><math>\beta \pm \text{s.e.}</math></b>	<b><math>z</math></b>	<b><math>P</math></b>
Intercept	-0.67 ± 1.13	-0.59	0.55
<b>Motivation group<sup>a</sup></b>	<b>-0.11 ± 0.53</b>	<b>-1.99</b>	<b>&gt;0.05</b>
Personality <sup>b</sup>	0.23 ± 0.48	0.47	0.64
Sex <sup>c</sup>	-0.41 ± 0.54	0.76	0.45
Latency to touch the device	<0.001 ± <0.001	0.36	0.72
Inhibition score	<0.01 ± <0.01	0.66	0.51



948 <sup>a</sup> low (reference level is high), <sup>b</sup> slow (reference level is fast), <sup>c</sup> male (reference level is female)

949 Figure 1. Routes of progression through the multi-access problem-solving experiment. To  
950 quantify their previous experience and propensity to pull sticks, individuals were initially  
951 presented with an opaque tube with a lever. They had four trials (30 mins each) in which to  
952 pull the lever. Once they solved the task once, they were classified as having previous  
953 experience solving a lever task. All birds progressed to the multi-access problem solving task  
954 where birds were presented with the transparent experimental device in which three access  
955 points were functional. Each bird had to solve the task using the same access point three  
956 times, before moving onto the next phase, where the previously solved access point was  
957 fused, leaving the remaining functional access points. This process was repeated for the other  
958 two access points. At any point of the testing, if a bird failed over three consecutive trials,  
959 their participation in the experiment ended. Dashed arrows indicate there is an alternative  
960 progression to complete the experiment.

961 Figure 2. The multi-access problem-solving device given to birds in their home cage. The  
962 apparatus has three different access types to retrieve the food reward inside; a lever, a swing  
963 door, and a string.

964 Figure 3 The effect of previous experience (whether there was an access point available that  
965 the bird had solved previously including the opaque device) on accuracy (the number of  
966 touches to functional parts of the device divided by all touches to the device per trial) ( $\beta \pm$   
967 s.e. =  $0.33 \pm 0.04$ ,  $t = 7.40$ ,  $P < 0.001$ , table 2). Note a previously solved access point could  
968 still be available as birds had to solve each access point three times before it was fused.

969 Figure 4 The frequency of access point solved, grouped by access point, ordered by trial  
970 number, as indicated in the legend.

971

972 Figure 5 Effect of motivation (high or low) on the number of different access points solved,  
973 ( $\beta \pm \text{s.e.} = -1.04 \pm 0.37$ ,  $z = -2.80$ ,  $P = 0.005$ , Table 4). Smaller points represent individual  
974 birds (which have been jittered along the x-axis and rendered partially transparent to reduce  
975 overlap; as a result, any remaining overlap results in darker points). The large black points  
976 represents the mean number of access points solved in each motivation group; the t-bars  
977 represent standard error.

978 Figure 6 Schematic of the study's results, with the four dependant variables aligned in the  
979 centre, arrows indicate influence of explanatory variables. Dashed arrows indicate a non-  
980 significant tendency, no arrows refer to non-significant relationship. Note that no test was  
981 performed between previous experience and accuracy, and innovativeness.

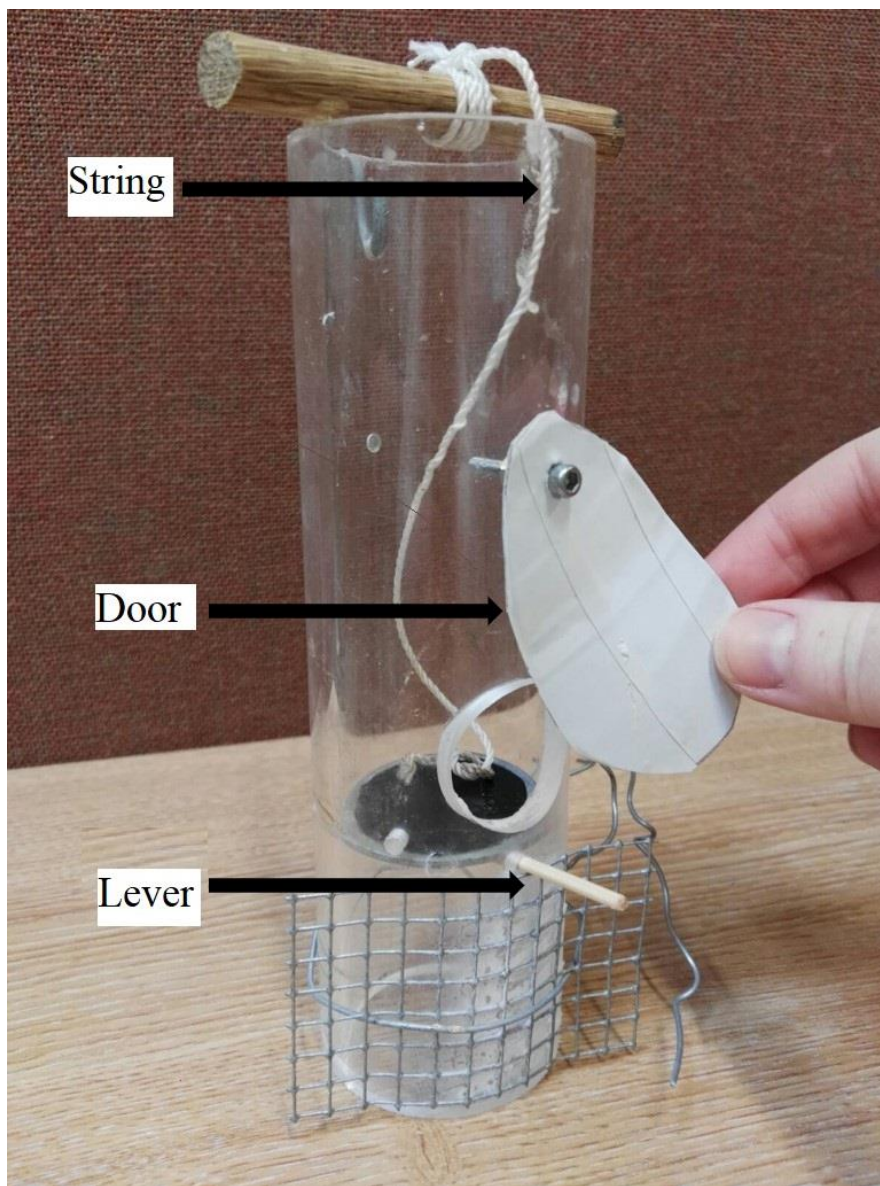
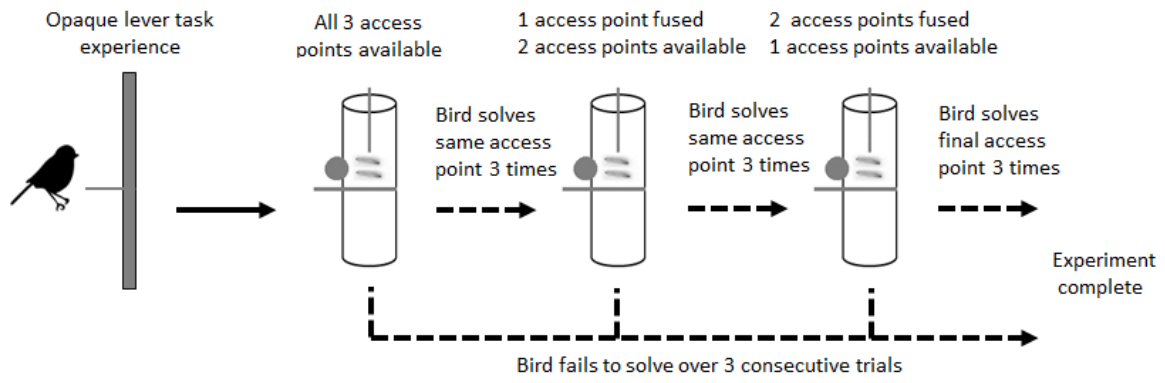
982 Figure A1 Latency to touch the device decreased significantly with trial number. Data is log  
983 transformed in this figure for scale purposes.

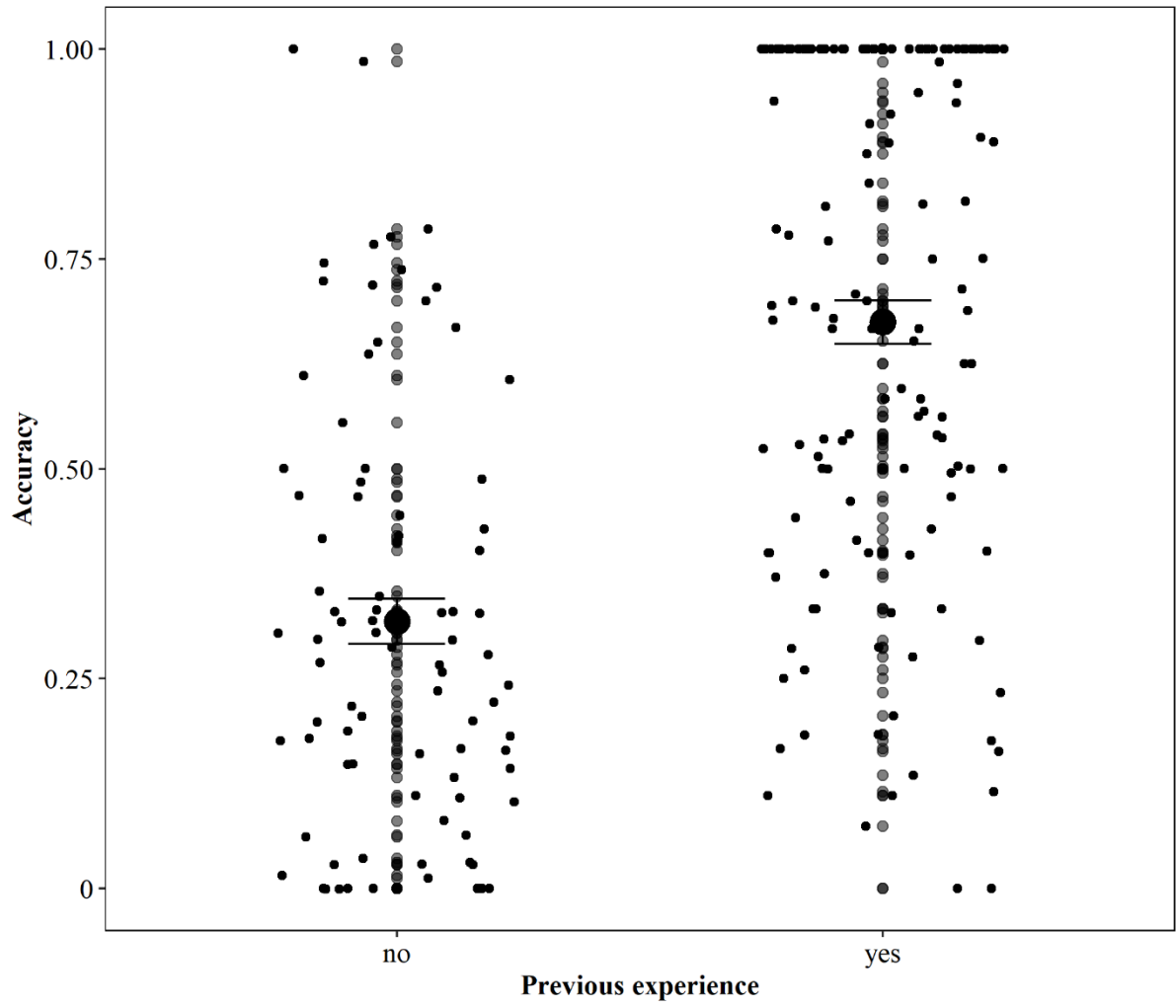
984

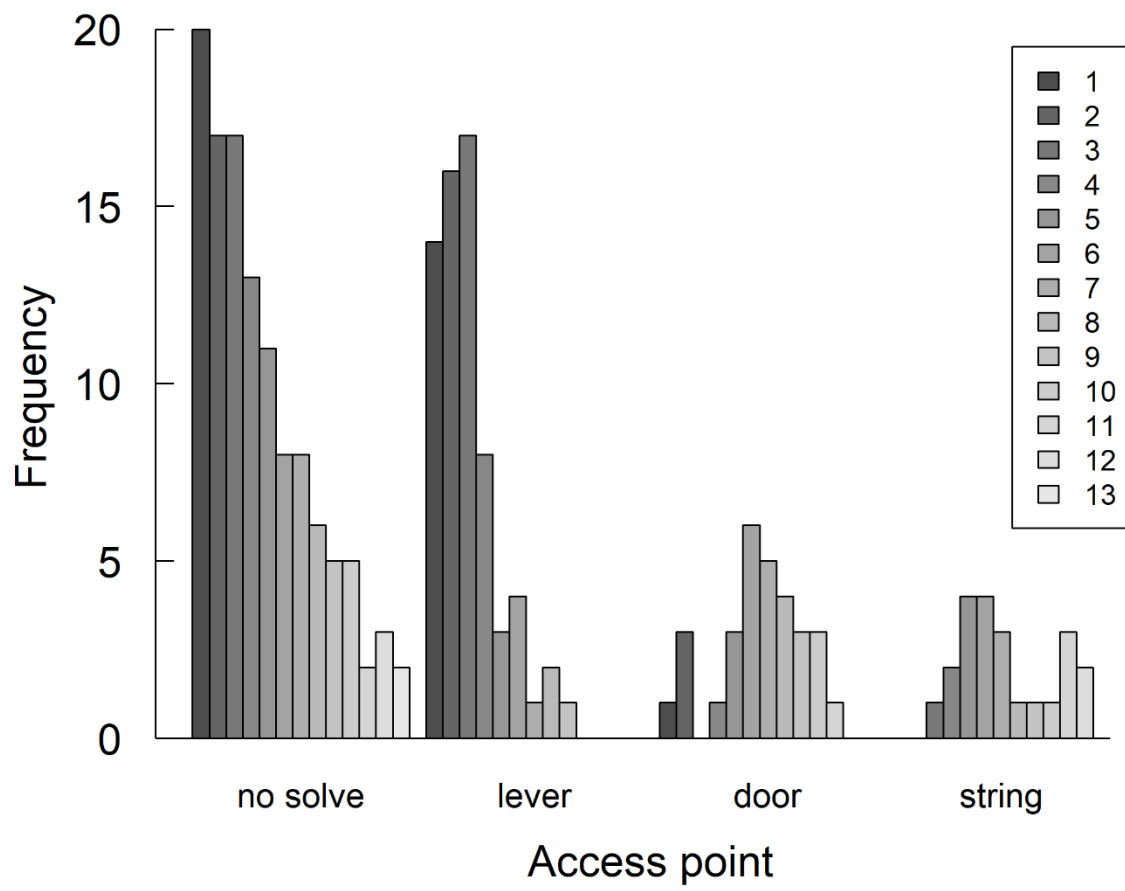
## ACKNOWLEDGEMENTS

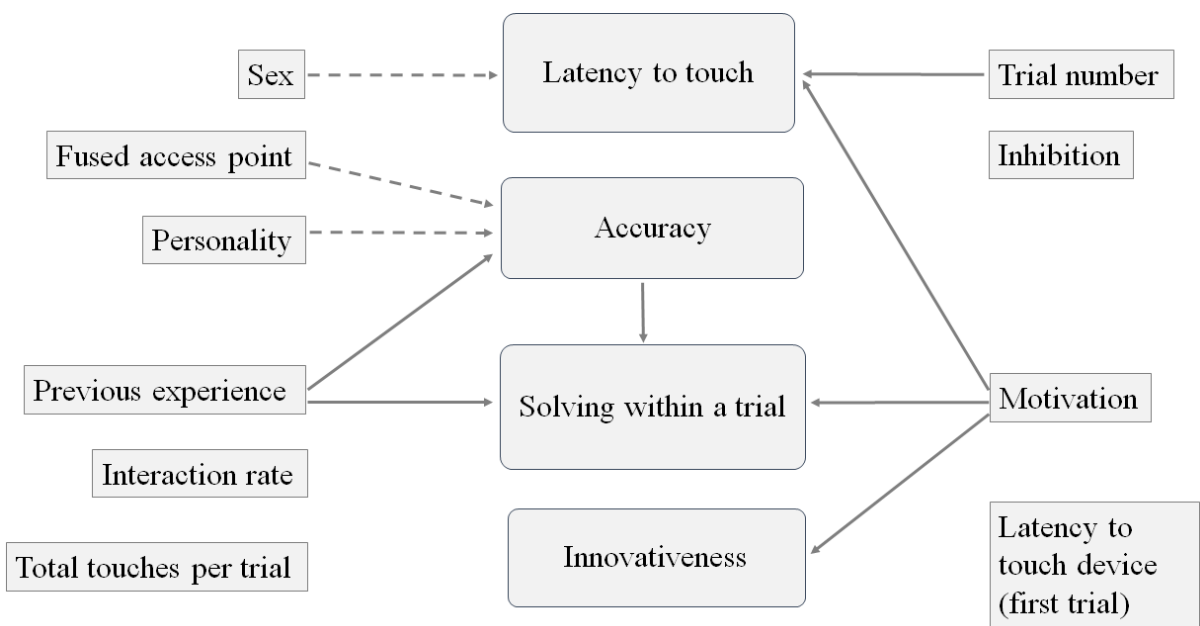
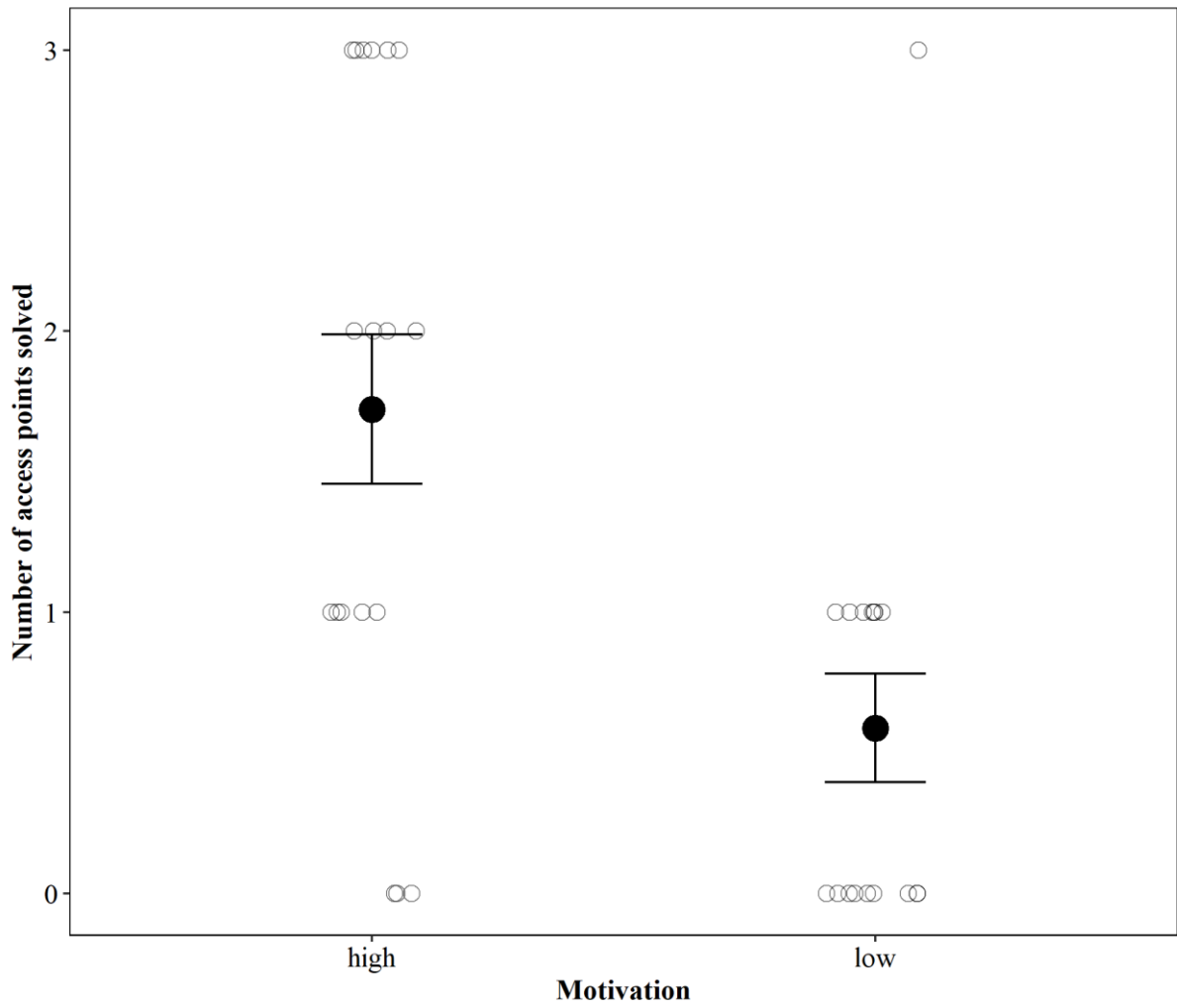
We would like to thank Anouk de Plaa, Ruben de Wit, Marylou Aaldering and Franca Kropman for taking care of the birds and helping with the experiments. We thank Rémi Pitiot for coding videos for our inter-observer reliability analysis, Mark Whittingham for advice on the analysis and Anthony Caravaggi for his assistance in generating the graphs. We thank Lucy Aplin and four anonymous referees for their constructive comments and suggestions during the review process. Funding for ACC and GLD from the European Research Council under the European Union's Horizon 2020 Programme (FP7/2007-2013)/ERC Consolidator Grant "EVOLECOCOG" Project No. 617509, awarded to JLQ, and by a Science Foundation Ireland ERC Support Grant 14/ERC/B3118 to JLQ.

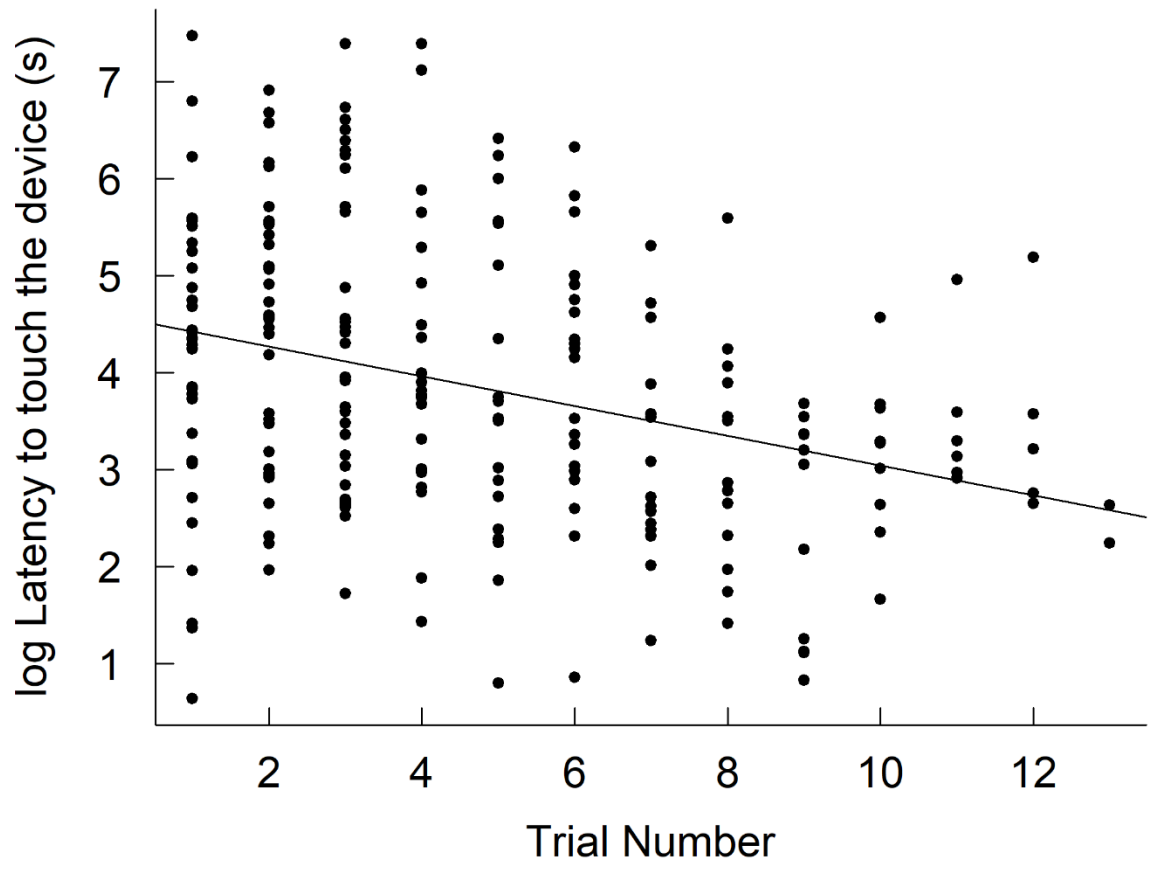
Figure















[Click here to access/download](#)

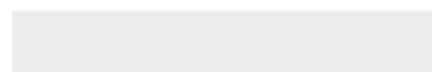
**Related material**

[Innovation\\_MS\\_master\\_data\\_per\\_trial.csv](#)





Click here to access/download  
**Related material**  
RCode\_Innovation\_MS.R



*Ethical Note*

We performed the experiment in accordance with the Association for the Study of Animal Behaviour guidelines. All experiments were approved by an ethical committee (DEC-KNAW licence no. NIOO 14.12 to KVO) and daily health checks were carried out to ensure the birds' welfare. Birds returned to the stock population after the behavioural experiments.