




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

Amy C. Cooke <sup>a, \*</sup> , Gabrielle L. Davidson <sup>a, b</sup> , Kees van Oers <sup>c</sup> , John L. Quinn <sup>a, d, \*</sup>

<sup>a</sup> School of Biological, Earth and Environmental Sciences, University College Cork, Cork, Ireland

<sup>b</sup> Department of Psychology, University of Cambridge, Cambridge, U.K.

<sup>c</sup> Department of Animal Ecology, Netherlands Institute of Ecology (NIOO-KNAW), Wageningen, the Netherlands

<sup>d</sup> Environmental Research Institute, University College Cork, Cork, Ireland

### ARTICLE INFO

#### Article history:

Received 26 March 2020

Initial acceptance 8 June 2020

Final acceptance 1 December 2020

Available online 4 March 2021

MS. number: 20-00198R

#### Keywords:

accuracy  
cognitive repeatability  
inhibitory control  
innovation  
motivation  
*Parus major*  
personality  
problem solving  
pseudorepeatability

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 use of a known behaviour 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, *Parus major*, and of their overall innovativeness across tasks. We gave animals originating from lines bidirectionally selected for fast or slow early exploratory behaviour, a multiaccess problem-solving device. Diverse motor skills and behavioural flexibility were required to solve all three different access points sequentially over trials. Food-deprived, highly motivated birds had shorter latency to touch the device, were more likely to solve an access point within a trial, and solved a greater diversity of them, than their less motivated counterparts. Solving success increased with accuracy when interacting with the device (proportion of touches to functional components of the device compared to all touches to the device per trial), 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: (1) pseudorepeatable effects (upward bias) linked to hunger-induced motivation, (2) repeatable differences in accuracy when interacting with devices, and (3) a feedback loop caused by experience gained over successive trials. Our results highlight the challenges of characterizing consistent individual differences in behaviour generally and support the idea that complex sources of variation play an important role in problem-solving performance.

© 2021 The Author(s). Published by Elsevier Ltd on behalf of The Association for the Study of Animal Behaviour. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Acquiring resources in changing environments is a major challenge faced by animals and a key determinant of fitness. Innovation, the generation of a novel behaviour or use of a known behaviour in a novel context, most commonly achieved through a problem-solving process, is one mechanism that a wide range of animals use to meet this challenge (Seed & Mayer, 2017). Comparative analysis has provided evidence for selection acting on innovativeness across species, because it helps animals find new food sources, or adapt to new environments and seasonal changes (Daniels,

Fanelli, Gilbert, & Benson-Amram, 2019; Lefebvre, Reader, & Sol, 2004; Reader, 2003; Reader & Laland, 2002; Sol, Lefebvre, & Rodríguez-Teijeiro, 2005; Webster & Lefebvre, 2001). Furthermore there is growing evidence of a link between innovativeness and fitness within populations (Cauchard, Boogert, Lefebvre, Dubois, & Doligez, 2013; Cole, Morand-Ferron, Hinks, & Quinn, 2012; Preiszner et al., 2017), and that innovation enables invasive or urbanized species to make use of novel resources (Daniels et al., 2019; Griffin & Diquelou, 2015; Griffin, Diquelou, & Perea, 2014). Although the underlying proximate causes of individual variation in innovativeness are diverse (for example, infection by parasites, Dunn, Cole, & Quinn, 2011; social factors, Thornton & Samson, 2012; natal environment effects, Kotrschal & Taborsky, 2010), repeatability analyses suggest differences between individuals are

\* Corresponding authors.

E-mail addresses: [cookeam@tcd.ie](mailto:cookeam@tcd.ie), [j.quinn@ucc.ie](mailto:j.quinn@ucc.ie) (A. C. Cooke), [cookeam@tcd.ie](mailto:cookeam@tcd.ie), [j.quinn@ucc.ie](mailto:j.quinn@ucc.ie) (J. L. Quinn).

consistent, pointing to intrinsic, potentially additive genetic, sources of variation (Cauchoix et al., 2018; Cole, Cram, & Quinn, 2011; Morand-Ferron, Cole, Rawles, & Quinn, 2011). One of the major challenges in the field is that innovativeness is a composite trait driven by a range of disparate behavioural processes that selection may act on independently and that may explain consistent differences in performance between individuals. These processes include cognition and motivation, as well as personality traits like exploration, persistence and neophobia (Griffin & Guez, 2014; Lermite, Peneaux, & Griffin, 2017; Seed & Call, 2010; Taylor, Hunt, Medina, & Gray, 2009). Thus, a key objective is to determine which processes drive innovativeness and explain the consistent individual differences observed.

Innovativeness correlates with relative brain size across species (Overington, Morand-Ferron, Boogert, & Lefebvre, 2009), suggesting an important role for cognitive mechanisms (Benson-Amram, Dantzer, Stricker, Swanson, & Holekamp, 2016; Lefebvre, Whittle, Lascaris, & Finkelstein, 1997; Reader & Laland, 2002; Sol, Duncan, Blackburn, Cassey, & Lefebvre, 2005). A number of cognitive mechanisms have been proposed to underlie innovative problem solving, including causal reasoning, insight, associative learning and inhibitory control (Barrett, Stanton, & Benson-Amram, 2019; Benson-Amram & Holekamp, 2012). Additionally, the ability to draw on previous experience aids individuals in reaching a solution (Cauchard et al., 2013; Ebel & Call, 2018; Griffin et al., 2014; Sol, Griffin, & Bartomeus, 2012), as well as ensuring the new behaviour becomes established in an individual's repertoire (Fragaszy et al., 2013). Finally, in situations where opportunities for innovation may be common, for example among invasive or urbanized species (Daniels et al., 2019; Griffin & Diquelou, 2015; Griffin et al., 2014), the ability to innovate frequently may be essential but dependent on inhibitory control (Daniels et al., 2019), whereby individuals must inhibit a previously rewarding behaviour that is no longer rewarding, in order to innovate further (Pecora et al., 2017). However, cognitive processes alone cannot explain why some individuals innovate more than others (van Horik & Madden, 2016), especially when other behavioural traits such as exploration and persistence also lead to innovation (Daniels et al., 2019; Ebel & Call, 2018; Overington, Cauchard, Côté, & Lefebvre, 2011).

Personality, defined as within-individual behavioural consistency across time and contexts (Réale, Reader, Sol, McDougall, & Dingemanse, 2007), provides a framework for exploring constraints on behavioural plasticity (Dall, Houston, & McNamara, 2004) and individual problem-solving performance (Hopper et al., 2014; Morton, Lee, & Buchanan-Smith, 2013). Personality traits have attracted particular attention because they predict individual variation in a wide range of behavioural traits (Aplin, Farine, Mann, & Sheldon, 2014; Cole & Quinn, 2012, 2014). Studies in the wild (Dingemanse, Both, Drent, van Oers, & van Noordwijk, 2002; Highcock & Carter, 2014) and in the laboratory (David, Auclair, & Cézilly, 2012; van Oers & Naguib, 2013) show that the personality trait 'early-life exploratory behaviour' (more specifically in this case, repeatable differences in the reaction to both a novel environment and objects; Drent, van Oers, & van Noordwijk, 2003) can influence how individuals retrieve information from their environment (Smit & van Oers, 2019), how quickly they solve problems (Hopper et al., 2014), and the degree of behavioural flexibility shown (Coppens, de Boer Sietse, & Koolhaas Jaap, 2010). In particular, fast-exploring (hereafter 'fast') individuals may be quicker to interact with or solve tasks (Benson-Amram & Holekamp, 2012; Trompf & Brown, 2014) but show less behavioural plasticity (Amy, van Oers, & Naguib, 2012; Jolles, Briggs, Araya-Ajoy, & Boogert, 2019; Logan, 2016c). Slow-exploring (hereafter 'slow') individuals tend to be the opposite (Coppens de Boer Sietse, & Koolhaas Jaap, 2010; Ducatez, Audet, & Lefebvre,

2015; Johnson-Ulrich, Johnson-Ulrich, & Holekamp, 2018; Sol et al., 2012; Zandberg, Quinn, Naguib, & van Oers, 2017). Additionally, neophobia (the fear of novel food, objects or places; Greenberg & Mettke-Hofmann, 2001) can constrain both the latency to approach a novel object and engagement in tasks. For example, individual hyenas, *Crocuta crocuta*, that showed greater persistence, activity or lower neophobia were faster to solve a problem (Johnson-Ulrich et al., 2018). However, the evolutionary significance of links between innovation and personality traits, as defined in Réale et al. (2007), is often unclear because the genetic basis for the personality variation is usually unknown (Cole et al., 2011), except in those few cases where personality-selective breeding lines have been used (Drent et al., 2003; van Oers Drent, de Goede, & van Noordwijk, 2004; van Oers, de Jong, van Noordwijk, Kempenaers, & Drent, 2005). Moreover, the role of other personality traits at different stages of innovative problem solving (e.g. interacting with a problem, solving a problem and ceasing to perform outdated solutions) and its interactions with other factors such as stress and motivation remain largely unexplored. Individual differences may be especially pronounced under stress (Suomi, 2004), but this has scarcely been tested. Note that although all behavioural variation can be defined as personality in a statistical sense (e.g. Dingemanse & Dochtermann, 2013), here we follow Réale et al. (2007) who focused on five kinds of behavioural traits, including exploration behaviour, that inherently capture variation in many other behavioural traits.

Motivation is expected to be an important driver of innovative behaviour (Laland & Reader, 1999; Sol et al., 2012) and to affect all stages of innovation. The 'necessity drives innovation' hypothesis states that innovative behaviours commonly occur when individuals are in need (Reader & Laland, 2003), that is, when they are motivated (Laland & Reader, 1999). For example, subdominant or juvenile individuals are often assumed to be more likely to innovate because they are less competitive when foraging (Morand-Ferron et al., 2011; Thornton & Samson, 2012). The rarely tested assumption in these studies is that hunger acts as the motivating factor driving innovation. In animal behaviour studies, food deprivation is commonly applied to ensure trial participation (Birch, 1945; Overington et al., 2011; Sol et al., 2012), or when attempting to control for confounding effects of motivation (Ebel & Call, 2018; van Horik & Madden, 2016). However, the extent to which motivation may influence innovative problem-solving behaviour at an individual level has scarcely been examined explicitly (Griffin & Guez, 2014).

Here we explore behavioural processes that are predicted to cause variation during sequential innovative problem solving, using second- and third-generation birds selected for personality. Selection lines are a powerful means to investigate inherent effects of personality on problem-solving performance as opposed to simple phenotype–phenotype correlations. We used a device that incorporated three different extractive foraging access points to provide a more complete measure of individual performance. The solutions relied on different motor skills, thus limiting the effects of individual motor skill bias, and previous motor skill experience carrying over to solving new access points. We examined variation in three different behavioural assays involved in innovative problem solving: (1) latency to touch the novel apparatus; (2) accuracy when interacting with any access point on the device; (3) problem-solving success within each trial. Then we examined (4) the individual's overall innovativeness (the number of different access points solved at least once across all trials). We considered a range of potential explanatory factors for these different behavioural facets, including extrinsic motivation (hunger state, the only experimentally manipulated factor), inhibitory control, previous experience and personality (fast/slow selection lines). In line with

the theory that individual differences may be more pronounced under stress (Suomi, 2004), we investigated the interaction between motivation and personality, assuming that birds in the high-motivation (food-deprived) treatment were more stressed than those in the low-motivation treatment. Finally, to determine whether individual differences were consistent, we estimated repeatability for (1)–(3) and examined whether controlling for fixed effects modified our estimates of repeatability. Repeatability sets the upper limit of heritability and is fundamental in studies on the evolutionary ecology of innovation and behaviour generally. Although uncontrolled confounding effects can potentially lead to an underestimate of repeatability, more commonly they lead to overestimates (pseudorepeatability) and sometimes explain repeatability entirely (Catry, Ruxton, Ratcliffe, Hamer, & Furness, 1999; Dingemans & Dochtermann, 2013; Westneat, Hatch, Wetzel, & Ensminger, 2011).

We predicted that: (1) birds in the high-motivation treatment group would have a reduced latency to touch the device, show increased accuracy (i.e. a high proportion of interactions with functional components, rather than nonfunctional components of the device), be more likely to solve an access point and solve more of them; (2) fast explorers would have a shorter latency to touch the device and lower accuracy when interacting with the device than slow explorers, but they may have a higher likelihood of solving due to higher exploration of the device; (3) previous experience would enable innovation, by causing a decrease in latency to touch the device, an increase in accuracy when interacting with it, and an increased likelihood of solving; (4) likelihood of solving in a trial would increase with accuracy (i.e. with higher frequency of interactions with functional components); and (5) birds with higher inhibition ability would be more likely to adjust their behaviour to solve multiple access points.

## METHODS

All experiments were carried out at the Netherlands Institute of Ecology (NIOO-KNAW), on 36 captive-bred great tits, *Parus major*. All birds included in the study were adult (2 years or older). Seventeen birds were not related to each other, five had one sibling and 14 shared more than one sibling; we assume relatedness between individuals had no bearing on the results. Birds were housed individually in standard cages (0.9 × 0.5 m and 0.5 m high) containing three perches and a water bath. Birds were in auditory contact but were visually isolated to prevent social learning. All birds had ad libitum access to water and a maintenance diet (ground beef heart, commercial egg food, fruit and calcium) unless otherwise stated. One bird did not participate in any of the experiments and was thus excluded from any analysis.

### Personality

Birds came from the second and third generation of bidirectionally phenotypically selected great tits, based on personality for 'fast exploration' (fast,  $N = 18$ ) and 'slow exploration' (slow,  $N = 18$ ). The measure of 'exploration' used during the selection process was a combination of two novel object tests where the latency to touch a novel object was recorded (e.g. a pink panther toy or an AA battery taped to a wooden stick), and one novel environment test where birds were released into a room and the latency to land on the fourth out of five artificial trees was recorded (for further details on selection and personality lines see Drent et al., 2003). The birds in the final selection lines used here underwent these same assays after fledging to confirm their personality type. As the specific aim of this study was to investigate the effects of artificially selected personality lines on problem solving

and because the bird behaviour matched their selected personality type, we analysed personality according to their selection history only (i.e. fast or slow selection lines).

### Motivation

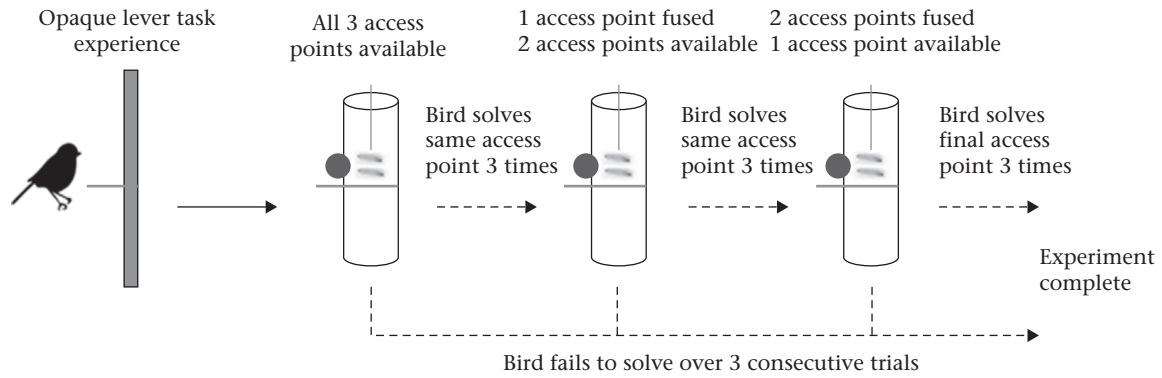
Individuals were randomly assigned to one of two motivation treatment groups for the duration of the experiment based on hunger state. The low-motivation group consisted of sated individuals, given full access to maintenance diet up to the start of the trial. Additionally, to ensure they were sated, they were given three wax moth, *Achroia grisella*, larvae 30 min before trials began, and invariably ate them all. The high-motivation group consisted of food-deprived birds, which had all sources of food removed from their cage for 1 h before the trial (Hämäläinen, Rowland, Mappes, & Thorogood, 2019). For welfare purposes, all birds had access to water during the trials. Motivation treatment was spread across the selection lines in four categories: high motivation, fast (female  $N = 4$  and male  $N = 5$ ); high motivation, slow (female  $N = 5$  and male  $N = 4$ ); low motivation, fast (female  $N = 5$  and male  $N = 4$ ); low motivation, slow (female  $N = 5$  and male  $N = 3$ ).

### Lever-pulling Propensity

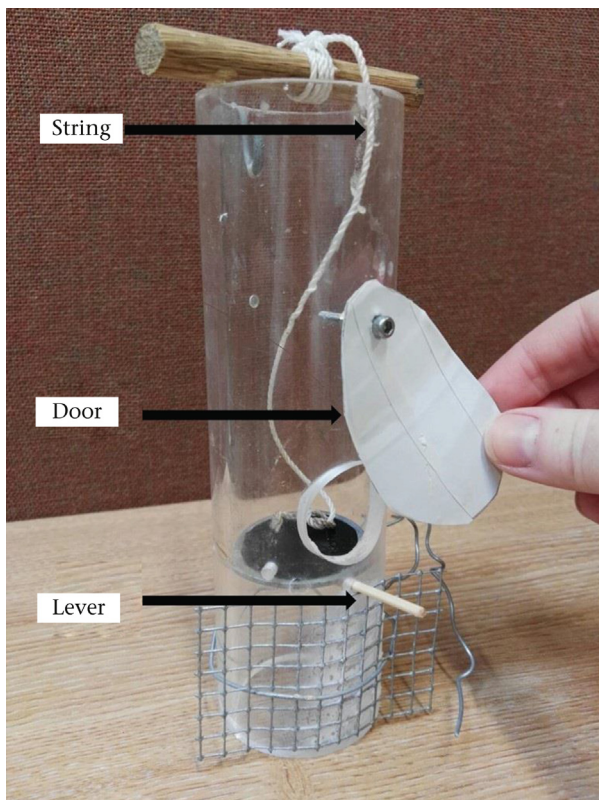
All trials described here and in the following section were carried out in the birds' individual home cages, under natural winter diurnal light cycles. To establish whether the birds had a pre-existing tendency to lever pull (von Bayern, Heathcote, Rutz, & Kacelnik, 2009), and because some birds may have had previous experience lever pulling in previous experiments while others had not, we measured lever-pulling propensity prior to testing them on the multiaccess device. We presented all birds with an opaque PVC rectangular tube containing a lever-supported platform with half a wax moth larva (Zandberg et al., 2017). We used an opaque device to test whether birds had a propensity to pull a stick, independent of a visual food reward cue, because the previous device that had been used was also opaque (Zandberg et al., 2017), and because we did not want the birds to have experience with the main innovation test device beforehand. All birds were given up to four trials (30 min per trial) to obtain the food reward, by pulling the lever horizontally causing the platform and reward to drop. Individuals that solved this opaque task at least once were classified (in the main analysis on the multiaccess problem-solving task described below) as having previous experience with solving the lever-pulling task. All birds progressed to the multiaccess task irrespective of their performance in this opaque device (Fig. 1).

### Multiaccess Problem-Solving Task

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



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



**Figure 2.** The multiaccess problem-solving device given to birds in their home cage. The apparatus had three different access types to retrieve the food reward inside: a lever, a swing door and a string.

Experiments were scheduled evenly across mornings and afternoons for both treatments and personality lines. In each trial, subjects were presented with the device and given 30 min to solve any of the access points. Birds were given two trials per day, back-to-back, without being fed between trials. Following their second trial, their maintenance diet was returned until testing the following day. The experiment ended when they had solved all

three access points three times, or when they had failed to solve any over three consecutive trials (total number of trials 3–13). Once an individual solved the same access point across three separate trials, that access point (door, lever or string) was fused, mimicking natural depletion of that food source, which meant that solving that access point was no longer possible, although it remained present and visible. We allowed birds to solve each access point three times to increase the chance that the behaviour became fixed in their repertoire. To solve a novel solution, they would need to behave flexibly, which we predicted would be guided by inhibitory control. Great tits from selection lines in this facility readily participate in experiments, so we assumed the three trials were sufficient to allow them to overcome any neophobic response.

All trials were recorded using a Panasonic HC-V250EB-K camera mounted on a tripod, covered in camouflage tape and positioned 1 m from the cage. Videos were analysed using Behavioural Observation Research Interactive Software (BORIS; Friard & Gamba, 2016). Observers were blind to the personality assigned to the birds but were aware of the motivation treatment group. Ten per cent of videos were coded by a second person. Interrater reliability was assessed using a Kendall's tau correlation test for agreement on the following measures: total number of touches to the device per trial,  $P < 0.001$ ; touches to functional access points on the device per trial,  $P < 0.01$ ; touches to anything other than functional access points on the device per trial,  $P < 0.005$ .

#### *Inhibition Task*

To generate an independent estimate of each individual's motor inhibition, we used a classical detour-reaching task (Beran, 2015; Boogert, Anderson, Peters, Searcy, & Nowicki, 2011; Rothbart & Posner, 1985), which tests to what extent the birds could control the prepotent response of pecking straight towards a food reward visible within a transparent Perspex tube. To pass the test, birds had to obtain the reward by accessing it through the opening on the side (Thorndike, 1911). The detour task was performed on a subset of 20 birds, prior to the problem-solving task (number of days between end of the detour-reaching task and first test day on the multiaccess device: mean  $\pm$  SE =  $11 \pm 0.46$ , minimum = 8, maximum = 12) to control for carryover experience with the transparent Perspex. Birds were not food deprived before this task.

There were three phases to this task: habituation, training and test phases. Birds participated in one phase per day, with progression through the phases occurring over consecutive days (duration of testing: mean  $\pm$  SE = 1.64  $\pm$  0.18 days, minimum = 1, maximum = 4). In the habituation and training phases, the Perspex tube was opaque (covered with black tape). To familiarize the birds with the apparatus, a wax moth larva was placed at the opening edge of the tube. Birds passed the habituation phase when they had eaten the reward three consecutive times. During the training phase, individuals had to obtain the food reward located in the centre of the opaque tube without touching any other part of the device. Training was completed when this was done successfully during four of five consecutive trials, ensuring the birds had the motor skills and experience necessary to move around the tube to successfully obtain the larva. During the test phase, the food reward was placed in the centre of a transparent tube. Birds had to remove the food reward without pecking on any other part of the device to complete the trial successfully. Inhibitory control scores were quantified as the number of trials it took individuals to complete four of five consecutive trials correctly. All trials were a maximum of 3 min each and observed remotely by livestreaming to a mobile phone using a Wi-Fi-enabled SJCAM SJ4000 camera (Shenzhen Zhencheng Technology, Shenzhen, China).

#### Ethical Note

We performed the experiment in accordance with the ASAB/ABS guidelines. All experiments were approved by an ethical committee (DEC-KNAW licence no. NIOO 14.12 to K.V.O.) and daily health checks were carried out to ensure the birds' welfare. Birds were returned to the stock population after the experiment.

#### Statistical Analysis

We tested whether multiple factors influenced different response variables at different stages of sequential innovative problem-solving performance: (1) latency to touch, (2) accuracy, (3) likelihood of solving and (4) innovativeness. Separate analyses were conducted using R Studio (R Studio Team, 2019) on each of the four phases described above (1–4), and we repeated these models on the subset of birds ( $N = 22$ ) that completed the inhibition task. For touch latency and accuracy, we conducted general linear mixed models (GLMMs) using the nlme package (Pinheiro, Bates, DebRoy, Sarkar, & R Core Team, 2019) fitted with a normal distribution; for likelihood of solving we ran a GLMM using the lme4 package (Bates, Mächler, Bolker, & Walker, 2015) fitted with a binomial distribution; and for innovativeness we ran a general linear model (GLM) with Poisson distribution (see Table A1 for a full list of variables and their definitions). In line with Whittingham, Stephens, Bradbury, and Freckleton (2006), we retained all variables of biological significance in the initial models to test specific hypotheses. For model selection, we used Akaike's information criterion (AIC) to measure goodness of fit (reported in table legends) and likelihood ratio tests to determine which model explained more variance. We compared full models (with the interaction between motivation and personality) to null models, and then compared full models to reduced models (i.e. without the interaction between motivation and personality). We dropped the interaction term from the model if the likelihood ratio test was nonsignificant ( $\alpha = 0.05$ ). To confirm that this hypothesis testing approach did not lead to a Type 2 error due to overfitting, we further reduced each model to the minimum adequate model using backwards reduction (see Tables A2–A5).

**Table 1**

Full model outputs from GLMM with factors affecting latency to touch the device per trial

| Fixed effects                    | $\beta \pm$ SE                     | $t$          | $P$              |
|----------------------------------|------------------------------------|--------------|------------------|
| Intercept                        | 4.27 $\pm$ 0.35                    | 12.69        | <0.001           |
| Motivation group <sup>a</sup>    | <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         | 0.504            |
| Trial number                     | <b>-0.09 <math>\pm</math> 0.03</b> | <b>-3.36</b> | <b>0.001</b>     |
| Sex <sup>c</sup>                 | -0.53 $\pm$ 0.30                   | -1.76        | 0.089            |
| Previous experience <sup>d</sup> | -0.19 $\pm$ 0.17                   | -1.15        | 0.25             |

$N = 226$ ,  $df = 190$ ,  $R^2$  (marginal = 0.26, conditional = 0.49, AIC = 739.83). Random effect included bird identity (0.72; 95% confidence interval, CI = 0.51, 1.02) and the residual variance (1.08; 95% CI = 0.97, 1.19). Significant results ( $P < 0.05$ ) are highlighted in bold.

<sup>a</sup> Low (reference level is high).

<sup>b</sup> Slow (reference level is fast).

<sup>c</sup> Male (reference level is female).

<sup>d</sup> Yes (reference level is no).

We checked that all models met assumptions (homogeneity, normality of residuals and collinearity of explanatory variables) using the DHARMA package in R (Hartig, 2020). We calculated confidence intervals (CI) for the random factor and residuals in each model using the package nlme in R (Pinheiro & Bates, 2006). In the legend of each table, we report marginal  $R^2$  (defined as the proportion of variance in the dependent variable that is explained by the fixed factors only), and conditional  $R^2$  (defined as the proportion of variance in the dependent variable that is explained by the fixed and random factors), or pseudo  $R^2$  (the marginal  $R^2$  of a Poisson GLM, which does not include random factors). We tested the following 4 models.

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

(2) Accuracy was defined as the number of touches to a functioning access point divided by the total number of touches to any part of the device per trial. Fixed effects included interaction rate (total number of touches to any part of the device per min, per trial), motivation group, trial number, selected personality lines, previous experience of any functioning (but not fused) access point (including previous experience of lever-pulling propensity on opaque device), fused trial (where any of the solutions were fused and therefore unavailable, as a fused access point may decrease accuracy) and sex. Individual bird identity was included as a random effect as subjects completed multiple trials.

(3) To test which factors predicted solving within each trial (binary;  $N = 224$ ), we included the following fixed effects: accuracy, previous experience, motivation group, personality, sex, trial number and fused trial. Individual bird identity was included as a random term. To limit overparameterization in the model, we did not include latency to touch in this analysis (but see analysis on number of different solves).

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

access points solved across all trials, we did not include variables that are trial specific (i.e. previous experience and accuracy).

Finally, we determined individual repeatability of the response variables in each of the first three questions above (latency to touch the device, accuracy and solving within a trial), using the rptR package, estimating repeatability (intraclass correlation) and CIs from Gaussian, binary, proportion and Poisson data (Stoffel, Nakagawa, & Schielzeth, 2017). We report unadjusted and adjusted repeatability, to encompass repeatability before and after controlling for influential fixed effects (Cauchoix et al., 2018). Unadjusted repeatability measures the between-individual variation in a given behaviour, while adjusted repeatability controls for fixed effects that could influence individual behaviour, because they explain either between- or within-individual components of variation. For both adjusted and unadjusted repeatability, we included individual identity as a random effect. Data and R code are included in the Supplementary material.

## RESULTS

### Latency to Touch the Multiaccess Device

Latency to touch the multiaccess device decreased over consecutive trials (Table 1, Fig. A1). The high-motivation group took less time to touch the device than the low-motivation group. Latency to touch the device did not differ between the personality selection lines. There was a nonsignificant trend for sex, suggesting that males took less time to touch the device than females. The variance of the random effect (individual bird identity) and the residual are indicated in Table 1. There was no effect of previous experience. The interaction between motivation and personality was not significant ( $\beta \pm SE = -0.47 \pm 0.60$ ,  $t = -0.78$ ,  $P = 0.44$ ). Inhibitory score had no effect on latency to touch the device (see Table A6).

### Accuracy

Birds were more accurate if they had previous experience solving any functioning access point, including solving the opaque device before the main experiment (Table 2, Fig. 3). There was a nonsignificant trend for slow birds being more accurate than fast birds. Birds tended to be less accurate in trials where there was a fused access point. The variance of the random effect (individual bird identity) and residual are indicated in Table 2. There was no effect of motivation group, interaction rate, sex or trial number on accuracy. The interaction between motivation and personality was

**Table 2**  
Full model outputs from GLMM with factors affecting accuracy per trial

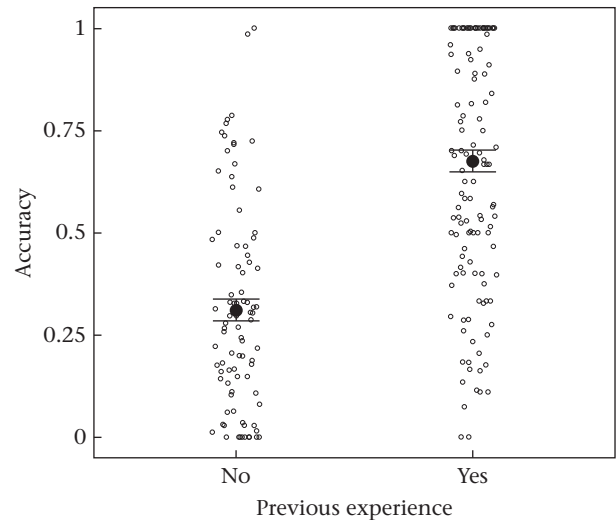
| Fixed                            | $\beta \pm SE$     | $t$         | $P$              |
|----------------------------------|--------------------|-------------|------------------|
| Intercept                        | 0.29 ± 0.08        | 3.44        | <0.001           |
| Interaction rate                 | 0.00 ± 0.00        | 0.43        | 0.670            |
| Previous experience <sup>a</sup> | <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            |

$N = 222$ ,  $df = 182$ ,  $R^2$  (marginal = 0.33, conditional = 0.56, AIC = 209.4). Random effect included bird identity (0.24; 95% confidence interval, CI = 0.22, 0.27) and the residual variance (0.15; 95% CI = 0.10, 0.23). Significant result ( $P < 0.05$ ) is highlighted in bold.

<sup>a</sup> Yes (reference level is no).

<sup>b</sup> Slow (reference level is fast).

<sup>c</sup> Low (reference level high).

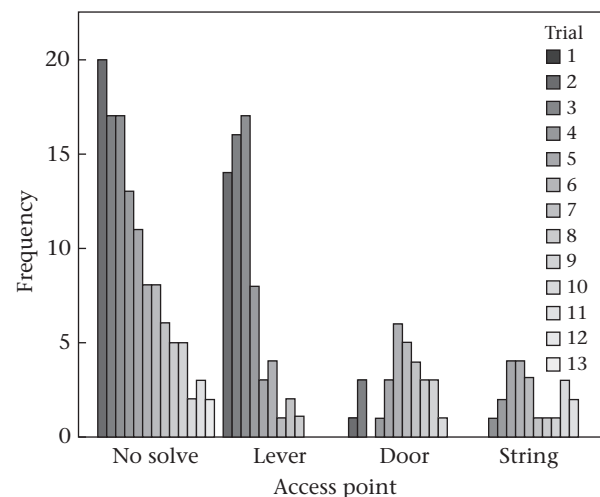


**Figure 3.** The effect of previous experience (whether there was an access point available that the bird had solved previously including the opaque device) on accuracy (the number of touches to functional parts of the device divided by all touches to the device per trial; see Table 2). Note that a previously solved access point could still be available as birds had to solve each access point three times before it was fused, thus making it unavailable. Smaller points represent individual birds (which have been jittered along the X axis to reduce overlap; as a result, any remaining overlap results in darker points). The large point represents the mean and the error bars represent SEs.

not significant ( $\beta \pm SE = -0.06 \pm 0.15$ ,  $t = -0.39$ ,  $P = 0.70$ ). Inhibition was unrelated to accuracy (see Table A7).

### Solving Within a Trial

Nineteen of the 35 birds pulled the lever on the opaque device. Of the 35 birds that participated in the multiaccess task, 12 solved one access point, four solved two access points, seven solved all three access points and 12 did not solve any (Fig. 4). Three birds solved three different access points over three consecutive trials while the device was fully operational (all access points functioning). One bird solved two access points in one trial, solving the string and then the lever in their fourth trial. We include both solutions as separate observations in our analysis. The lever was



**Figure 4.** The frequency of access points solved, grouped by access point, ordered by trial number, as indicated.

**Table 3**

Full model outputs from GLMM with factors affecting solving within trials

| Fixed effects                     | $\beta \pm SE$                     | $z$          | $P$              |
|-----------------------------------|------------------------------------|--------------|------------------|
| Intercept                         | $-1.93 \pm 0.68$                   | 2.86         | 0.004            |
| Accuracy                          | <b><math>3.79 \pm 0.78</math></b>  | <b>4.90</b>  | <b>&lt;0.001</b> |
| Previous experience <sup>a</sup>  | <b><math>1.26 \pm 0.47</math></b>  | <b>2.66</b>  | <b>0.008</b>     |
| Motivation group <sup>b</sup>     | <b><math>-1.79 \pm 0.50</math></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 <sup>e</sup>          | $-0.18 \pm 0.73$                   | -0.25        | 0.806            |
| Total number of touches per trial | $-0.00 \pm 0.00$                   | -1.47        | 0.141            |

$N = 224$ ,  $df = 216$ ,  $R^2$  (marginal = 0.51, conditional = 0.52, AIC = 55.85). Random effect included bird identity (0.37; 95% confidence interval, CI = 0.41, 0.33) and the residual variance (0.13; 95% CI = 0.10, 0.27). Significant results ( $P < 0.05$ ) are highlighted in bold.

<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> Male (reference level is female).

<sup>e</sup> Yes (reference level is no).

solved by 23 different birds, the door was solved by 10 birds and the string by seven birds.

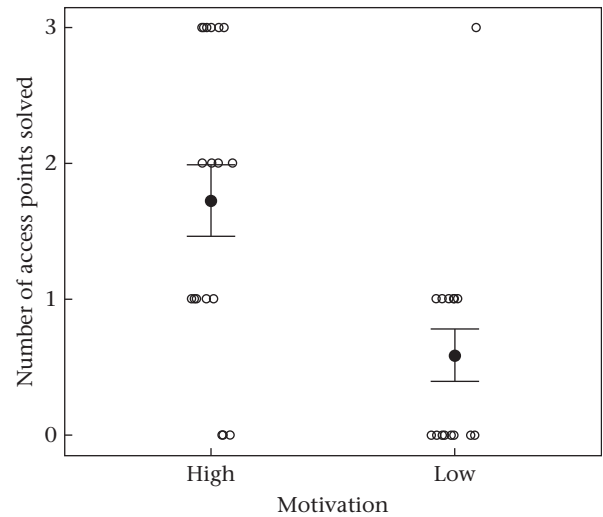
Food-deprived birds were more likely to solve an access point (Table 3). Higher accuracy and previous experience also predicted solving likelihood within a trial. The variance of the individual identity random effect and the residual are indicated in Table 3. There was no effect of personality, sex, trial number, whether it was a fused trial or not, total number of touches to device per trial or inhibition (see Table A8). The interaction between motivation and personality was not significant ( $\beta \pm SE = -0.46 \pm 0.90$ ,  $t = -0.51$ ,  $P = 0.61$ ). Follow-up post hoc analysis, using a Fisher's exact test, revealed a correlation trend between lever pulling on the opaque and multiaccess device ( $P = 0.07$ ). Further analysis, investigating the order in which the multiaccess device was solved, using a Fisher's exact test, showed that the lever was more likely to be solved first ( $P < 0.001$ ), while there was no difference between the string or door (Fig. 4).

#### Innovativeness: Number of Access Points Solved

Highly motivated birds solved more novel access points than low-motivated birds (Table 4, Fig. 5). There was no effect of personality, sex, latency to touch the device in the first trial only (Table 4) or inhibition (Table A9). The interaction between personality and motivation was nonsignificant ( $\beta \pm SE = -0.86 \pm 0.79$ ,  $z = -1.09$ ,  $P = 0.28$ ).

#### Repeatability

Latency to touch the device was repeatable but repeatability decreased when adjusted for significant fixed effects (Table 5). Accuracy was also repeatable but increased when adjusted for



**Figure 5.** Effect of motivation (high or low) on the number of different access points solved (see Table 4). Smaller points represent individual birds (which have been jittered along the X axis to reduce overlap; as a result, any remaining overlap results in darker points). The black points represent the mean number of access points solved in each motivation group and the error bars represent SEs.

significant effects. Solving performance within a trial was also repeatable but repeatability disappeared entirely when adjusted for all significant fixed effects. To further investigate which factors were reducing the individual repeatability between the unadjusted and adjusted  $R$  values for solving access points within a trial, we removed each fixed effect individually and reran the repeatability model (see Table 5). Adjusted repeatability changed only when a significant fixed effect was excluded. Adjusted repeatability without accuracy was significant, and without motivation, while adjusted repeatability without previous experience only approached significance. There was no change in adjusted repeatability for any factor that did not affect solving performance.

## DISCUSSION

Our study sought to explore factors that drive individual variation and repeatability at various stages of innovative problem-solving performance (Fig. 6). We showed that hunger-induced motivation affected multiple problem-solving stages, that previous experience influenced accuracy, and that hunger, accuracy and previous experience influenced problem-solving success. Personality and inhibitory control had little or no effect. Solvers of the opaque lever-pulling device tended to solve the lever on the multiaccess device. Furthermore, birds were more likely to solve the lever first, but showed no preference between the door and string. All traits were significantly repeatable; however, the repeatability

**Table 4**

Full model outputs from GLM with factors affecting the number of different access points solved by an individual

| Explanatory variables                             | $\beta \pm SE$                     | $z$          | $P$          |
|---|------------------------------------|--------------|--------------|
| Intercept   | $<0.001 \pm <0.001$                | 1.84         | 0.066        |
| Motivation group <sup>a</sup>                     | <b><math>-1.04 \pm 0.37</math></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        |

$N = 35$ ,  $df = 24$ , pseudo- $R^2 = 0.24$ , AIC = 99.60. Significant result ( $P < 0.05$ ) is highlighted in bold.

<sup>a</sup> Low (reference level is high).

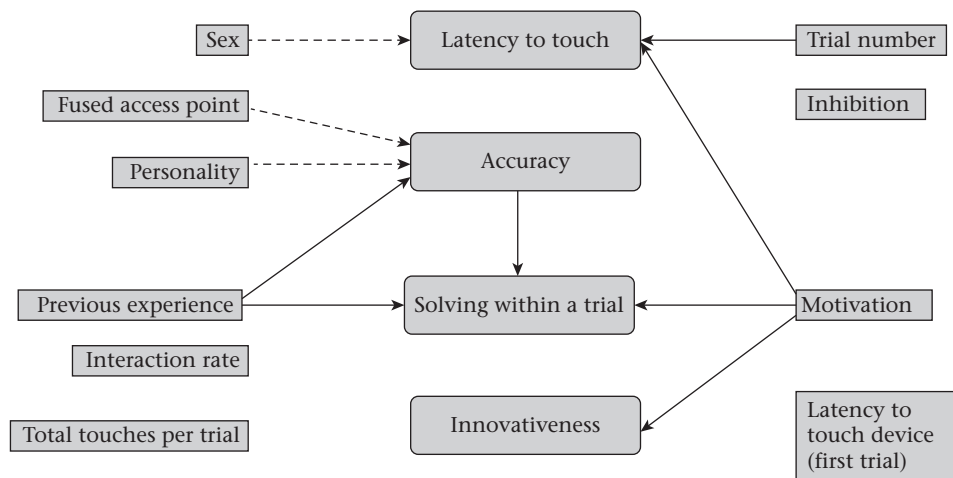
<sup>b</sup> Slow (reference level is fast).

<sup>c</sup> Male (reference level is female).

**Table 5**  
Repeatability (adjusted and unadjusted) estimates for the three main components of problem-solving behaviour during the experiment

| Model                                | Repeatability                         | R           | Confidence interval (lower, upper) | P                |
|--------------------------------------|---------------------------------------|-------------|------------------------------------|------------------|
| Latency to touch the device          | Adjusted                              | <b>0.32</b> | <b>0.15, 0.45</b>                  | <b>&lt;0.001</b> |
|                                      | Unadjusted                            | <b>0.48</b> | <b>0.30, 0.61</b>                  | <b>&lt;0.001</b> |
| Accuracy                             | Adjusted                              | <b>0.45</b> | <b>0.17, 0.52</b>                  | <b>&lt;0.001</b> |
|                                      | Unadjusted                            | <b>0.23</b> | <b>0.08, 0.38</b>                  | <b>0.001</b>     |
| Solving access points within a trial | Adjusted                              | 0.03        | 0, 0.17                            | 0.40             |
|                                      | Unadjusted                            | <b>0.31</b> | <b>0.1, 0.46</b>                   | <b>&lt;0.001</b> |
|                                      | Adjusted: without accuracy            | <b>0.18</b> | <b>0, 0.33</b>                     | <b>0.044</b>     |
|                                      | Adjusted: without motivation group    | <b>0.21</b> | <b>0, 0.35</b>                     | <b>0.017</b>     |
|                                      | Adjusted: without previous experience | 0.14        | 0, 0.29                            | 0.076            |

Unadjusted values are from mixed models with only individual as a random effect. Adjusted values also include significant fixed effects for each of superscripts a, b and c as shown in Tables 1, 2 and 3, respectively. In addition, for superscript c adjusted repeatabilities are also shown when single fixed effects were removed. Significant results ( $P < 0.05$ ) are highlighted in bold.



**Figure 6.** Schematic of the study's main results, with the four dependent variables aligned in the centre; arrows indicate influence of explanatory variables (left or right side). Dashed arrows indicate a nonsignificant tendency, no arrows refer to nonsignificant relationship. Note that no test was performed between previous experience and accuracy, and innovativeness because the two former variables were measured per trial, while the latter measure accrued across all trials.

of problem solving was explained entirely by motivation, accuracy and experience.

#### Motivation Drives Innovation

Although motivation is often viewed as a confounding variable, if considered at all when examining mechanisms underlying problem-solving tasks (reviewed in Griffin & Guez, 2014), it also underpins the 'necessity drives innovation' hypothesis (Reader & Laland, 2003). In support of this hypothesis, motivation was the major driver of an individual's latency to touch the device, to solve the same access point repeatedly and to innovate multiple times in our experimental set-up. Previous studies reported that task engagement increased with increased food deprivation, thus facilitating problem solving (Griffin et al., 2014; Sol et al., 2012), but motivation itself did not predict problem solving (Griffin & Guez, 2014; van Horik & Madden, 2016). Likewise, the relationship between problem solving and motivation, as measured by body weight or body condition, is inconclusive: at times an effect is present (Laland & Reader, 1999; Mateos-Gonzalez, Quesada, & Senar, 2011) and other times not (Cole et al., 2011; Thornton & Samson, 2012; but see Griffin & Guez, 2014, for a full review). This variability in results across studies may be in part due to

differences in how motivation is defined and how problem solving is measured. Our results emphasize the importance of controlling for motivation and standardizing the length of time animals are food deprived in captive experiments, as well as acknowledging that not knowing an animal's motivational state may be a weakness of cognitive experiments conducted in the wild. Nevertheless, controlling for motivational effects generally is unlikely to be straightforward (Auersperg, Gajdon, & Bayern, 2012; Griffin & Guez, 2014; Morand-Ferron et al., 2016; Morand-Ferron & Quinn, 2011), not least because whether food deprivation removes, or just changes, individual variation remains unclear.

#### Personality

Considerable evidence suggests that personality traits defined by Réale et al. (2007) influence individual problem-solving performance (Greenberg, 2003; Johnson-Ulrich et al., 2018; Sol, Griffin, Bartomeus, & Boyce, 2011). However, in our study, personality selection lines with known genetic provenance for object neophobia and novel environment exploration did not predict latency to touch the device; nor did they predict problem-solving behaviour, in terms of success within trials, or the number of different innovations reached. We predicted the effects of hunger-induced



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

### *Inhibitory Control*

Inhibitory control is an integral part of behavioural flexibility (MacLean et al., 2014; Manrique, Völter, & Call, 2013), both of which are beneficial for problem solving, allowing animals to overcome outdated information. Contrary to our predictions, individuals that exhibited high inhibitory control were no more likely to generate a novel solution to the task than those with low inhibitory control, even when the reward contingencies changed (i.e. when an access point was fused), a time when behavioural flexibility is required. This lack of correlation may be because changing one's behaviour is necessary but not sufficient to solve a problem (Logan, 2016a, 2016b). Moreover, the validity of the detour-reaching task as a test for inhibitory control remains under debate because performance does not necessarily correlate with other tasks that aim to measure inhibitory control, or because previous experience of transparency and persistence may influence performance (Kabadayi, Bobrowicz, & Osvath, 2018; van Horik et al., 2018). Neither Johnson-Ulrich et al. (2018) nor Daniels et al. (2019) found a correlation between problem solving and inhibitory control, even when inhibitory control was measured using an alternative paradigm to the detour-reaching task. Thus we conclude that there is no case for motor inhibition affecting behavioural flexibility in the context of problem solving, but it remains possible that it reflects other facets of behavioural flexibility (reviewed in Bari & Robbins, 2013).

### *Previous Experience*

Birds with previous experience of having solved the opaque lever device, or indeed any of the three access points during the main trials, were more accurate and had higher solving success in subsequent trials. Furthermore, performance improved with experience over repeated problem-solving attempts with regard to that particular solving method, perhaps owing to instrumental conditioning. Thus, attributing an individual's cognitive performance to how quickly it solves a problem, or its ability to solve multiple novel problems, may be a function of its previous experience (Rowe & Healy, 2014; Sih & Del Giudice, 2012). We acknowledge the constraints in controlling for all experiences animals may have had with features of an experimental apparatus, especially if based on simple generalizable rules. Nevertheless, tasks could be designed such that they include multiple access points that vary in modality (e.g. smell and touch: sensory versus

motor), in the appearance of the specific materials they use (e.g. white plastic versus black plastic) and/or in the required motor skills as we have attempted to do here (Auersperg, Bayern, Gajdon, Huber, & Kacelnik, 2011; Griffin & Guez, 2014; Manrique et al., 2013; Overington et al., 2009). This paradigm may facilitate the testing of true innovations that are not confounded by previous experience, or alternatively, to explicitly test what kinds of experiences facilitate future innovations.

### *Repeatability, Pseudorepeatability and Positive Feedback*

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

The significance of these findings is tied to the nature of the specific factor involved. First, in the case of hunger, designed to manipulate motivation, each individual only experienced one of two treatments, a potentially reversible effect, suggesting that the component of the unadjusted between-individual difference explained by hunger was inflated, resulting in pseudorepeatability. Although some sources of motivation are probably permanent, either through a permanent environment (Wilson, 2018) or intrinsic motivation (Ebel & Call, 2018; Gajdon, Lichtegger, & Huber, 2014; Polizzi di Sorrentino et al., 2014; Taffoni et al., 2014), this pseudo measure demonstrates that failure to control for motivation caused by temporary factors can inflate the intrinsic between-individual differences that researchers are attempting to characterize; that is, those differences that are caused by permanent environment or intrinsic effects. Second, accuracy explained some of the between-individual variation, suggesting that the mechanisms underlying accurate interaction with the device vary consistently between individuals themselves, and explain some of the between-individual differences in the problem-solving performance. It appears likely these mechanisms are intrinsic rather than reversible since motivation is controlled for in these analyses. Third, experience also caused some of the between-individual differences in problem-solving performance, and since experience is not reversible, and by definition carries forward into the next stage of the sequential problem-solving process, this suggests a positive feedback loop driving consistency between individual differences in problem-solving behaviour. Although the role of feedback loops in driving differences in individual behaviour is well known (Dall et al., 2004; Sih et al., 2015), and examples of positive feedbacks are common in nature (Kishida et al., 2011), to our knowledge none have explained consistent between-individual differences. In this case we assume the feedback caused by experience leads to a permanent effect, although it remains possible that individuals eventually forget the experience.

Our results highlight the challenges of characterizing consistent individual variation in sequential problem-solving performance as a measure of overall innovativeness. More generally, they provide a demonstration of how between-individual differences in innovation can be explained by inflated estimates of within-individual variation in motivation, inflated between-individual variation in accuracy, and by feedback loops involving previous experience. Much of the focus in studies on the evolutionary ecology of behaviour in general has been on the evolutionary processes that drive intrinsic individual variation. Our results support the idea that complex sources of variation underlying single traits are likely to make predicting the selective consequences of this variation challenging.

## Acknowledgments

We 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 interobserver 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 A.C.C. and G.L.D. came from the European Research Council under the European Union's Horizon 2020 Programme (FP7/2007–2013)/ERC Consolidator Grant “EVOLECOG” Project No. 617509, awarded to J.L.Q., and from a Science Foundation Ireland ERC Support Grant 14/ERC/B3118 to J.L.Q.

## Supplementary Material

Supplementary material associated with this article can be found online at <https://doi.org/10.1016/j.anbehav.2021.01.024>.

## References

- Amy, M., van Oers, K., & Naguib, M. (2012). Worms under cover: Relationships between performance in learning tasks and personality in great tits (*Parus major*). *Animal Cognition*, 15(5), 763–770. <https://doi.org/10.1007/s10071-012-0500-3>
- Aplin, L. M., Farine, D. R., Mann, R. P., & Sheldon, B. C. (2014). Individual-level personality influences social foraging and collective behaviour in wild birds. *Proceedings of the Royal Society B: Biological Sciences*, 281(1789), 20140106. <https://doi.org/10.1098/rspb.2014.1016>
- Auersperg, A. M. I., Bayern, A. M. P. von, Gajdon, G. K., Huber, L., & Kacelnik, A. (2011). Flexibility in problem solving and tool use of kea and new caledonian crows in a multi access box paradigm. *PLoS One*, 6(6), Article e20231. <https://doi.org/10.1371/journal.pone.0020231>
- Auersperg, A. M. I., Gajdon, G. K., & Bayern, A. M. P. von (2012). A new approach to comparing problem solving, flexibility and innovation. *Communicative & Integrative Biology*, 5(2), 140–145. <https://doi.org/10.4161/cib.18787>
- Bari, A., & Robbins, T. W. (2013). Inhibition and impulsivity: Behavioral and neural basis of response control. *Progress in Neurobiology*, 108, 44–79. <https://doi.org/10.1016/j.pneurobio.2013.06.005>
- Barrett, L. P., Stanton, L. A., & Benson-Amram, S. (2019). The cognition of ‘nuisance’ species. *Animal Behaviour*, 147, 167–177. <https://doi.org/10.1016/j.anbehav.2018.05.005>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- von Bayern, A. M. P., Heathcote, R. J. P., Rutz, C., & Kacelnik, A. (2009). The role of experience in problem solving and innovative tool use in crows. *Current Biology*, 19(22), 1965–1968. <https://doi.org/10.1016/j.cub.2009.10.037>
- Benson-Amram, S., Dantzer, B., Stricker, G., Swanson, E. M., & Holekamp, K. E. (2016). Brain size predicts problem-solving ability in mammalian carnivores. *Proceedings of the National Academy of Sciences*, 113(9), 2532–2537. <https://doi.org/10.1073/pnas.1505913113>
- Benson-Amram, S., & Holekamp, K. E. (2012). Innovative problem solving by wild spotted hyenas. *Proceedings of the Royal Society B: Biological Sciences*, 279(1744), 4087–4095. <https://doi.org/10.1098/rspb.2012.1450>
- Beran, M. J. (2015). The comparative science of “self-control”: What are we talking about? *Frontiers in Psychology*, 6. <https://doi.org/10.3389/fpsyg.2015.00051>
- Birch, H. G. (1945). The role of motivational factors in insightful problem-solving. *Journal of Comparative Psychology*, 38(5), 295–317. <https://doi.org/10.1037/h0059937>
- Boogert, N. J., Anderson, R. C., Peters, S., Searcy, W. A., & Nowicki, S. (2011). Song repertoire size in male song sparrows correlates with detour reaching, but not with other cognitive measures. *Animal Behaviour*, 81(6), 1209–1216. <https://doi.org/10.1016/j.anbehav.2011.03.004>
- Catry, P., Ruxton, G. D., Ratcliffe, N., Hamer, K. C., & Furness, R. W. (1999). Short-lived repeatabilities in long-lived great skuas: Implications for the study of individual quality. *Oikos*, 84(3), 473–479. <https://doi.org/10.2307/3546426>
- Cauchard, L., Boogert, N. J., Lefebvre, L., Dubois, F., & Doligez, B. (2013). Problem-solving performance is correlated with reproductive success in a wild bird population. *Animal Behaviour*, 85(1), 19–26. <https://doi.org/10.1016/j.anbehav.2012.10.005>
- Cauchoix, M., Chow, P. K. Y., van Horik, J. O., Atance, C. M., Barbeau, E. J., Barragan-Jason, G., et al. (2018). The repeatability of cognitive performance: A meta-analysis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1756), 20170281. <https://doi.org/10.1098/rstb.2017.0281>
- Cole, E. F., Cram, D. L., & Quinn, J. L. (2011). Individual variation in spontaneous problem-solving performance among wild great tits. *Animal Behaviour*, 81(2), 491–498. <https://doi.org/10.1016/j.anbehav.2010.11.025>
- Cole, E. F., Morand-Ferron, J., Hinks, A. E., & Quinn, J. L. (2012). Cognitive ability influences reproductive life history variation in the wild. *Current Biology*, 22(19), 1808–1812. <https://doi.org/10.1016/j.cub.2012.07.051>
- Cole, E. F., & Quinn, J. L. (2012). Personality and problem-solving performance explain competitive ability in the wild. In *Proceedings of the Royal Society of London B: Biological Sciences*, Article rspb20111539. <https://doi.org/10.1098/rspb.2011.1539>
- Cole, E. F., & Quinn, J. L. (2014). Shy birds play it safe: Personality in captivity predicts risk responsiveness during reproduction in the wild. *Biology Letters*, 10(5), 20140178. <https://doi.org/10.1098/rsbl.2014.0178>
- Coppens, C. M., de Boer Sietse, F., & Koolhaas Jaap, M. (2010). Coping styles and behavioural flexibility: Towards underlying mechanisms. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1560), 4021–4028. <https://doi.org/10.1098/rstb.2010.0217>
- Dall, S. R. X., Houston, A. I., & McNamara, J. M. (2004). The behavioural ecology of personality: Consistent individual differences from an adaptive perspective. *Ecology Letters*, 7(8), 734–739. <https://doi.org/10.1111/j.1461-0248.2004.00618.x>
- Daniels, S. E., Fanelli, R. E., Gilbert, A., & Benson-Amram, S. (2019). Behavioural flexibility of a generalist carnivore. *Animal Cognition*, 22, 387–396. <https://doi.org/10.1007/s10071-019-01252-7>
- David, M., Auclair, Y., & Cézilly, F. (2012). Assessing short- and long-term repeatability and stability of personality in captive zebra finches using longitudinal data. *Ethology*, 118(10), 932–942. <https://doi.org/10.1111/j.1439-0310.2012.02085.x>
- Dingemans, N. J., Both, C., Drent, P. J., van Oers, K., & van Noordwijk, A. J. (2002). Repeatability and heritability of exploratory behaviour in great tits from the wild. *Animal Behaviour*, 64(6), 929–938. <https://doi.org/10.1006/anbe.2002.2006>
- Dingemans, N. J., & Dochtermann, N. A. (2013). Quantifying individual variation in behaviour: Mixed-effect modelling approaches. *Journal of Animal Ecology*, 82(1), 39–54. <https://doi.org/10.1111/1365-2656.12013>
- Drent, P. J., van Oers, K., & van Noordwijk, A. J. (2003). Realized heritability of personalities in the great tit (*Parus major*). *Proceedings of the Royal Society of London Series B Biological Sciences*, 270(1510), 45–51. <https://doi.org/10.1098/rspb.2002.2168>
- Ducatez, S., Audet, J. N., & Lefebvre, L. (2015). Problem-solving and learning in Carib grackles: Individuals show a consistent speed–accuracy trade-off. *Animal Cognition*, 18(2), 485–496. <https://doi.org/10.1007/s10071-014-0817-1>
- Dunn, J. C., Cole, E. F., & Quinn, J. L. (2011). Personality and parasites: Sex-dependent associations between avian malaria infection and multiple behavioural traits. *Behavioral Ecology and Sociobiology*, 65(7), 1459–1471. <https://doi.org/10.1007/s00265-011-1156-8>
- Ebel, S. J., & Call, J. (2018). The interplay of prior experience and motivation in great ape problem-solving (*Gorilla gorilla*, *Pan paniscus*, *Pan troglodytes*, and *Pongo abelii*). *Journal of Comparative Psychology*, 132(3), 294–305.
- Fragaszy, D. M., Biro, D., Eshchar, Y., Humle, T., Izar, P., Resende, B., et al. (2013). The fourth dimension of tool use: Temporally enduring artefacts aid primates learning to use tools. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368(1630), 20120410. <https://doi.org/10.1098/rstb.2012.0410>
- Friard, O., & Gamba, M. (2016). Boris: A free, versatile open-source event-logging software for video/audio coding and live observations. *Methods in Ecology and Evolution*, 7(11), 1325–1330. <https://doi.org/10.1111/2041-210X.12584>
- Gajdon, G. K., Lichtegger, M., & Huber, L. (2014). What a parrot’s mind adds to play: The urge to produce novelty fosters tool use acquisition in kea. *Open Journal of Animal Sciences*, 4(2), 51–58. <https://doi.org/10.4236/ojas.2014.42008>
- Greenberg, R. (2003). The role of neophobia and neophilia in the development of innovative behaviour of birds. In S. M. Reader, & K. N. Laland (Eds.), *Animal*

- innovation (pp. 175–196). Oxford, UK: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780198526223.003.0008>.
- Greenberg, R. S., & Mettke-Hofmann, C. (2001). Ecological aspects of neophobia and neophilia in birds. <http://repository.si.edu/handle/10088/8053>.
- Griffin, A. S., & Diquelou, M. C. (2015). Innovative problem solving in birds: A cross-species comparison of two highly successful passerines. *Animal Behaviour*, *100*, 84–94. <https://doi.org/10.1016/j.anbehav.2014.11.012>
- Griffin, A. S., Diquelou, M., & Perea, M. (2014). Innovative problem solving in birds: A key role of motor diversity. *Animal Behaviour*, *92*, 221–227. <https://doi.org/10.1016/j.anbehav.2014.04.009>
- Griffin, A. S., & Guez, D. (2014). Innovation and problem solving: A review of common mechanisms. *Behavioural Processes*, *109*, 121–134. <https://doi.org/10.1016/j.beproc.2014.08.027>
- Hämäläinen, L., Rowland, H. M., Mappes, J., & Thorogood, R. (2019). The effect of social information from live demonstrators compared to video playback on blue tit foraging decisions. *PeerJ*, *7*, Article e7998. <https://doi.org/10.7717/peerj.7998>
- Hartig, F. (2020). DHARMA: Residual diagnostics for hierarchical (multi-level/mixed) regression models (R package version 0.3.3.0). Computer software <https://CRAN.R-project.org/package=DHARMA>.
- Highcock, L., & Carter, A. J. (2014). Intraindividual variability of boldness is repeatable across contexts in a wild lizard. *PLoS One*, *9*(4), Article e95179. <https://doi.org/10.1371/journal.pone.0095179>
- Hopper, L. M., Price, S. A., Freeman, H. D., Lambeth, S. P., Schapiro, S. J., & Kendal, R. L. (2014). Influence of personality, age, sex, and estrous state on chimpanzee problem-solving success. *Animal Cognition*, *17*(4), 835–847. <https://doi.org/10.1007/s10071-013-0715-y>
- van Horik, J. O., Langley, E. J. G., Whiteside, M. A., Laker, P. R., Beardsworth, C. E., & Madden, J. R. (2018). Do detour tasks provide accurate assays of inhibitory control? *Proceedings of the Royal Society B: Biological Sciences*, *285*(1875), 20180150. <https://doi.org/10.1098/rspb.2018.0150>
- van Horik, J. O., & Madden, J. R. (2016). A problem with problem solving: Motivational traits, but not cognition, predict success on novel operant foraging tasks. *Animal Behaviour*, *114*, 189–198. <https://doi.org/10.1016/j.anbehav.2016.02.006>
- Johnson-Ulrich, L., Johnson-Ulrich, Z., & Holekamp, K. (2018). Proactive behavior, but not inhibitory control, predicts repeated innovation by spotted hyenas tested with a multi-access box. *Animal Cognition*, *21*(3), 379–392. <https://doi.org/10.1007/s10071-018-1174-2>
- Jolles, J. W., Briggs, H. D., Araya-Ajoy, Y. G., & Boogert, N. J. (2019). Personality, plasticity and predictability in sticklebacks: Bold fish are less plastic and more predictable than shy fish. *Animal Behaviour*, *154*, 193–202. <https://doi.org/10.1016/j.anbehav.2019.06.022>
- Kabadayi, C., Bobrowicz, K., & Osavath, M. (2018). The detour paradigm in animal cognition. *Animal Cognition*, *21*(1), 21–35. <https://doi.org/10.1007/s10071-017-1152-0>
- Kishida, O., Trussell, G. C., Ohno, A., Kuwano, S., Ikawa, T., & Nishimura, K. (2011). Predation risk suppresses the positive feedback between size structure and cannibalism. *Journal of Animal Ecology*, *80*(6), 1278–1287. <https://doi.org/10.1111/j.1365-2656.2011.01871.x>
- Kotrschal, A., & Taborsky, B. (2010). Environmental change enhances cognitive abilities in fish. *PLoS Biology*, *8*(4), Article e1000351. <https://doi.org/10.1371/journal.pbio.1000351>
- Laland, K. N., & Reader, S. M. (1999). Foraging innovation in the guppy. *Animal Behaviour*, *57*(2), 331–340. <https://doi.org/10.1006/anbe.1998.0967>
- Lefebvre, L., Reader, S. M., & Sol, D. (2004). Brains, innovations and evolution in birds and primates. *Brain, Behavior and Evolution*, *63*(4), 233–246. <https://doi.org/10.1159/000076784>
- Lefebvre, L., Whittle, P., Lascaris, E., & Finkelstein, A. (1997). Feeding innovations and forebrain size in birds. *Animal Behaviour*, *53*(3), 549–560. <https://doi.org/10.1006/anbe.1996.0330>
- Lermite, F., Peneaux, C., & Griffin, A. S. (2017). Personality and problem-solving in common mynas (*Acridotheres tristis*). *Behavioural Processes*, *134*, 87–94. <https://doi.org/10.1016/j.beproc.2016.09.013>
- Logan, C. J. (2016a). How far will a behaviourally flexible invasive bird go to innovate? *Royal Society Open Science*, *3*(6), 160247. <https://doi.org/10.1098/rsos.160247>
- Logan, C. J. (2016b). Behavioral flexibility and problem solving in an invasive bird. *PeerJ*, *4*, Article e1975. <https://doi.org/10.7717/peerj.1975>
- Logan, C. J. (2016c). Behavioral flexibility in an invasive bird is independent of other behaviors. *PeerJ*, *4*, e2215. <https://doi.org/10.7717/peerj.2215>
- MacLean, E. L., Hare, B., Nunn, C. L., Addessi, E., Amici, F., Anderson, R. C., et al. (2014). The evolution of self-control. *Proceedings of the National Academy of Sciences*, *111*(20), E2140–E2148. <https://doi.org/10.1073/pnas.1323533111>
- Manrique, H. M., Völter, C. J., & Call, J. (2013). Repeated innovation in great apes. *Animal Behaviour*, *85*(1), 195–202. <https://doi.org/10.1016/j.anbehav.2012.10.026>
- Mateos-Gonzalez, F., Quesada, J., & Senar, J. C. (2011). Sexy birds are superior at solving a foraging problem. *Biology Letters*, *7*(5), 668–669. <https://doi.org/10.1098/rsbl.2011.0163>
- Morand-Ferron, J., Cole, E. F., Rawles, J. E. C., & Quinn, J. L. (2011). Who are the innovators? A field experiment with 2 passerine species. *Behavioral Ecology*, *22*(6), 1241–1248. <https://doi.org/10.1093/beheco/arr120>
- Morand-Ferron, J., & Quinn, J. L. (2011). Larger groups of passerines are more efficient problem solvers in the wild. *Proceedings of the National Academy of Sciences*, *108*(38), 15898–15903. <https://doi.org/10.1073/pnas.1111560108>
- Morand-Ferron, J., Cole, E. F., & Quinn, J. L. (2016). Studying the evolutionary ecology of cognition in the wild: A review of practical and conceptual challenges. *Biological Reviews*, *91*(2), 367–389. <https://doi.org/10.1111/brv.12174>
- Morton, F. B., Lee, P. C., & Buchanan-Smith, H. M. (2013). Taking personality selection bias seriously in animal cognition research: A case study in capuchin monkeys (*Sapajus apella*). *Animal Cognition*, *16*(4), 677–684. <https://doi.org/10.1007/s10071-013-0603-5>
- van Oers, K., Drent, P. J., de Goede, P., & van Noordwijk, A. J. (2004). Realized heritability and repeatability of risk-taking behaviour in relation to avian personalities. *Proceedings of the Royal Society London Series B Biological Sciences*, *271*(1534), 65–73. <https://doi.org/10.1098/rspb.2003.2518>
- van Oers, K., de Jong, G., van Noordwijk, A. J., Kempnaers, B., & Drent, P. J. (2005). Contribution of genetics to the study of animal personalities: A review of case studies. *Behaviour*, *142*(9/10), 1185–1206.
- van Oers, K., & Naguib, M. (2013). Avian personality. In C. Carere, & D. Maestripieri (Eds.), *Animal personalities: Behavior, physiology, and evolution* (pp. 66–95). Chicago, IL: University of Chicago Press.
- Overington, S. E., Cauchard, L., Côté, K.-A., & Lefebvre, L. (2011). Innovative foraging behaviour in birds: What characterizes an innovator? *Behavioural Processes*, *87*(3), 274–285. <https://doi.org/10.1016/j.beproc.2011.06.002>
- Overington, S. E., Morand-Ferron, J., Boogert, N. J., & Lefebvre, L. (2009). Technical innovations drive the relationship between innovativeness and residual brain size in birds. *Animal Behaviour*, *78*(4), 1001–1010. <https://doi.org/10.1016/j.anbehav.2009.06.033>
- Pecora, G., Zoratto, F., Paoletti, M., Bellagamba, F., Paglieri, F., & Addessi, E. (2017). Executive function. In J. Vonk, & T. Shackelford (Eds.), *Encyclopedia of animal cognition and behavior* (pp. 1–5). New York, NY: Springer International Publishing. [https://doi.org/10.1007/978-3-319-47829-6\\_1048-1](https://doi.org/10.1007/978-3-319-47829-6_1048-1).
- Pinheiro, J., & Bates, D. (2006). *Mixed-Effects Models in S and S-PLUS*. New York, NY: Springer Science & Business Media.
- Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team. (2019). nlme: Linear and nonlinear mixed effects models (3.1-141). Computer software <https://CRAN.R-project.org/package=nlme>.
- Polizzi di Sorrentino, E., Sabbatini, G., Truppa, V., Bordonali, A., Taffoni, F., Formica, D., et al. (2014). Exploration and learning in capuchin monkeys (*Sapajus* spp.): The role of action–outcome contingencies. *Animal Cognition*, *17*(5), 1081–1088. <https://doi.org/10.1007/s10071-014-0740-5>
- Preisner, B., Papp, S., Pipoly, I., Seress, G., Vincze, E., Liker, A., et al. (2017). Problem-solving performance and reproductive success of great tits in urban and forest habitats. *Animal Cognition*, *20*(1), 53–63. <https://doi.org/10.1007/s10071-016-1008-z>
- R Studio Team. (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.rstudio.org/>.
- Reader, S. (2003). Innovation and social learning: Individual variation and brain evolution. *Animal Biology*, *53*(2), 147–158. <https://doi.org/10.1163/157075603769700340>
- Reader, S. M., & Laland, K. N. (2002). Social intelligence, innovation, and enhanced brain size in primates. *Proceedings of the National Academy of Sciences*, *99*(7), 4436–4441. <https://doi.org/10.1073/pnas.062041299>
- Reader, S. M., & Laland, K. N. (Eds.). (2003). *Animal innovation*. Oxford, U.K.: Oxford University Press.
- Réale, D., Reader, S. M., Sol, D., McDougall, P. T., & Dingemanse, N. J. (2007). Integrating animal temperament within ecology and evolution. *Biological Reviews*, *82*(2), 291–318. <https://doi.org/10.1111/j.1469-185X.2007.00010.x>
- Rothbart, M. K., & Posner, M. I. (1985). Temperament and the development of self-regulation. In L. C. Hartlage, & C. F. Tiezrow (Eds.), *The neuropsychology of individual differences: A developmental perspective* (pp. 93–125). New York, NY: Plenum.
- Rowe, C., & Healy, S. D. (2014). Measuring variation in cognition. *Behavioral Ecology*, *25*(6), 1287–1292. <https://doi.org/10.1093/beheco/aru090>
- Seed, A. M., & Call, J. (2010). Problem-solving in tool-using and non-tool-using animals. In M. D. Breed, & J. Moore (Eds.), *Encyclopedia of animal behavior* (pp. 778–785). Cambridge, MA: Academic Press. <https://doi.org/10.1016/B978-0-08-045337-8.00090-5>.
- Seed, A. M., & Mayer, C. M. P. (2017). Problem solving. In *Perception, Learning, and Cognition: Vol. 2. APA Handbook of Comparative Psychology*. Washington, D.C.: American Psychological Association.
- Sih, A., & Del Giudice, M. (2012). Linking behavioural syndromes and cognition: A behavioural ecology perspective. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1603), 2762–2772. <https://doi.org/10.1098/rstb.2012.0216>
- Sih, A., Mathot, K. J., Moirón, M., Montiglio, P.-O., Wolf, M., & Dingemanse, N. J. (2015). Animal personality and state-behaviour feedbacks: A review and guide for empiricists. *Trends in Ecology & Evolution*, *30*(1), 50–60. <https://doi.org/10.1016/j.tree.2014.11.004>
- Smit, J. A. H., & van Oers, K. (2019). Personality types vary in their personal and social information use. *Animal Behaviour* <https://doi.org/10.1016/j.anbehav.2019.02.002>.
- Sol, D., Duncan, R. P., Blackburn, T. M., Cassey, P., & Lefebvre, L. (2005). Big brains, enhanced cognition, and response of birds to novel environments. *Proceedings of the National Academy of Sciences*, *102*(15), 5460–5465. <https://doi.org/10.1073/pnas.0408145102>
- Sol, D., Griffin, A. S., & Bartomeus, I. (2012). Consumer and motor innovation in the common myna: The role of motivation and emotional responses. *Animal Behaviour*, *83*(1), 179–188. <https://doi.org/10.1016/j.anbehav.2011.10.024>

- Sol, D., Griffin, A. S., Bartomeus, I., & Boyce, H. (2011). Exploring or avoiding novel food resources? The novelty conflict in an invasive bird. *PLoS One*, 6(5), Article e19535. <https://doi.org/10.1371/journal.pone.0019535>
- Sol, D., Lefebvre, L., & Rodríguez-Tejedor, J. D. (2005). Brain size, innovative propensity and migratory behaviour in temperate Palaearctic birds. *Proceedings of the Royal Society B: Biological Sciences*, 272(1571), 1433–1441. <https://doi.org/10.1098/rspb.2005.3099>
- Stoffel, M. A., Nakagawa, S., & Schielzeth, H. (2017). rptR: Repeatability estimation and variance decomposition by generalized linear mixed-effects models. *Methods in Ecology and Evolution*, 8(11), 1639–1644. <https://doi.org/10.1111/2041-210X.12797>
- Suomi, S. J. (2004). How gene-environment interactions shape biobehavioral development: Lessons from studies with rhesus monkeys. *Research in Human Development*, 1(3), 205–222. [https://doi.org/10.1207/s15427617rhd0103\\_5](https://doi.org/10.1207/s15427617rhd0103_5)
- Taffoni, F., Tamilia, E., Focaroli, V., Formica, D., Ricci, L., Di Pino, G., et al. (2014). Development of goal-directed action selection guided by intrinsic motivations: An experiment with children. *Experimental Brain Research*, 232(7), 2167–2177. <https://doi.org/10.1007/s00221-014-3907-z>
- Taylor, A. H., Hunt, G. R., Medina, F. S., & Gray, R. D. (2009). Do New Caledonian crows solve physical problems through causal reasoning? *Proceedings of the Royal Society B: Biological Sciences*, 276(1655), 247–254. <https://doi.org/10.1098/rspb.2008.1107>
- Thorndike, E. (1911). *Animal intelligence: Experimental studies*. London, U.K.: Macmillan.
- Thornton, A., & Samson, J. (2012). Innovative problem solving in wild meerkats. *Animal Behaviour*, 83(6), 1459–1468. <https://doi.org/10.1016/j.anbehav.2012.03.018>
- Trompf, L., & Brown, C. (2014). Personality affects learning and trade-offs between private and social information in guppies, *Poecilia reticulata*. *Animal Behaviour*, 88, 99–106. <https://doi.org/10.1016/j.anbehav.2013.11.022>
- Verbeek, M. E. M., Drent, P. J., & Wiepkema, P. R. (1994). Consistent individual differences in early exploratory behaviour of male great tits. *Animal Behaviour*, 48(5), 1113–1121. <https://doi.org/10.1006/anbe.1994.1344>
- Webster, S. J., & Lefebvre, L. (2001). Problem solving and neophobia in a columbiform–passeriform assemblage in Barbados. *Animal Behaviour*, 62(1), 23–32. <https://doi.org/10.1006/anbe.2000.1725>
- Westneat, D. F., Hatch, M. L., Wetzal, D. P., & Ensminger, A. L. (2011). Individual variation in parental care reaction norms: Integration of personality and plasticity. *The American Naturalist*, 178(5), 652–667. <https://doi.org/10.1086/662173>
- Whittingham, M. J., Stephens, P. A., Bradbury, R. B., & Freckleton, R. P. (2006). Why do we still use stepwise modelling in ecology and behaviour? *Journal of Animal Ecology*, 75(5), 1182–1189. <https://doi.org/10.1111/j.1365-2656.2006.01141.x>
- Wilson, A. J. (2018). How should we interpret estimates of individual repeatability? *Evolution Letters*, 2(1), 4–8. <https://doi.org/10.1002/evl3.40>
- Zandberg, L., Quinn, J. L., Naguib, M., & van Oers, K. (2017). Personality-dependent differences in problem-solving performance in a social context reflect foraging strategies. *Behavioural Processes*, 134, 95–102. <https://doi.org/10.1016/j.beproc.2016.09.007>

## Appendix

**Table A1**  
Variable names and definitions

| Variable                             | Definition  |
|--------------------------------------|---|
| Latency to touch                     | Latency from start of trial to touching the device per trial  |
| Personality                          | Second and third generation of bidirectionally selected personality lines for fast and slow exploring personality   |
| Motivation                           | Two motivation treatment groups, low motivation (sated birds, fed 3 wax moth larvae 30 min before trial commences) or high motivation (birds that were food deprived 1 h before trial)  |
| Trial number                         | Number of the trial   |
| Fused trial                          | When any access point was fused in the trial  |
| Bird identity                        | Identity code for each bird   |
| Number of access points solved       | Absolute number of different access points on the device solved by individual across all trials   |
| Sex                                  | Male or female  |
| Inhibition score                     | Number of trials until individual reaches criterion in detour-reaching task (criterion = 4/5 consecutive successful trials)   |
| Solving access points within a trial | Whether an individual has solved access points 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, nonfused access point available to solve that they had solved previously; for the lever in the multiaccess 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 of interacting with the device (time from first touch to last touch per trial)  |

**Table A2**

Minimum adequate model outputs from GLMM with factors affecting latency to touch the device per trial

| Fixed effects                 | $\beta \pm SE$                     | $t$          | $P$              |
|-------------------------------|------------------------------------|--------------|------------------|
| Intercept                     | 4.25 $\pm$ 0.27                    | 15.93        | <0.001           |
| Motivation group <sup>a</sup> | <b>1.19 <math>\pm</math> 0.29</b>  | <b>4.10</b>  | <b>&lt;0.001</b> |
| Trial number                  | <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             |

$N = 226$ ,  $df = 190$ ,  $R^2$  (marginal = 0.26, conditional = 0.47, AIC = 726.50). Random effect included bird identity (1.24; 95% confidence interval, CI = 1.08, 1.41) and the residual variance (0.51; 95% CI = 0.28, 0.95). Significant results ( $P < 0.05$ ) are highlighted in bold.

<sup>a</sup> Low (reference level is high).

<sup>b</sup> Male (reference level is female).

**Table A3**

Minimum adequate model outputs from GLMM with factors affecting accuracy per trial

| Fixed effects                    | $\beta \pm SE$                     | $t$          | $P$              |
|----------------------------------|------------------------------------|--------------|------------------|
| Intercept                        | 0.27 $\pm$ 0.05                    | 5.05         | <0.001           |
| Previous experience <sup>a</sup> | <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            |
| Fused trial                      | <b>-0.11 <math>\pm</math> 0.04</b> | <b>-2.55</b> | <b>0.012</b>     |

$N = 222$ ,  $df = 187$ ,  $R^2$  (marginal = 0.34, conditional = 0.54, AIC = 49.42). Random effect included bird identity (0.24; 95% confidence interval, CI = 0.20, 0.27) and the residual variance (0.19; 95% CI = 0.13, 0.30). Significant results ( $P < 0.05$ ) are highlighted in bold.

<sup>a</sup> Yes (reference level is no).

<sup>b</sup> Slow (reference level is fast).

**Table A4**

Minimum adequate model output from GLMM with factors affecting solving within trials

| Fixed effects                    | $\beta \pm SE$                     | $Z$          | $P$              |
|----------------------------------|------------------------------------|--------------|------------------|
| Intercept                        | -2.42 $\pm$ 0.50                   | -4.88        | <0.001           |
| Accuracy                         | <b>3.91 <math>\pm</math> 0.78</b>  | <b>4.99</b>  | <b>&lt;0.001</b> |
| Previous experience <sup>a</sup> | <b>1.52 <math>\pm</math> 0.42</b>  | <b>3.65</b>  | <b>&lt;0.001</b> |
| Motivation group <sup>b</sup>    | <b>-1.64 <math>\pm</math> 0.48</b> | <b>-3.36</b> | <b>&lt;0.001</b> |

$N = 224$ ,  $df = 223$ ,  $R^2$  (marginal = 0.55, conditional = 0.58, AIC = 205.70). Random effect included bird identity (0.37; 95% confidence interval, CI = 0.33, 0.41) and the residual variance (0.13; 95% CI = 0.09, 0.25). Significant results ( $P < 0.05$ ) are highlighted in bold.

<sup>a</sup> Yes (reference level is no).

<sup>b</sup> Low (reference level is high).

**Table A5**

Minimum adequate model outputs from GLM with factors affecting the number of different access points solved by an individual

| Explanatory variables         | $\beta \pm SE$      | <i>z</i>     | <i>P</i>         |
|-------------------------------|---------------------|--------------|------------------|
| Intercept                     | 0.54 ± 0.18         | 3.03         | <0.005           |
| Motivation group <sup>a</sup> | <b>-1.07 ± 0.36</b> | <b>-2.95</b> | <b>&lt;0.005</b> |

*N* = 35, *df* = 33, pseudo-*R*<sup>2</sup> = 0.23, AIC = 93.54. Significant result (*P* < 0.05) is highlighted in bold.

<sup>a</sup> Low (reference level is high).

**Table A6**

Full model outputs from GLMM with factors affecting latency to touch

| Fixed effects                    | $\beta \pm SE$      | <i>t</i>     | <i>P</i>        |
|----------------------------------|---------------------|--------------|-----------------|
| Intercept                        | 4.69 ± 0.90         | 5.22         | <0.001          |
| Motivation group <sup>a</sup>    | <b>1.13 ± 0.44</b>  | <b>2.58</b>  | <b>&lt;0.05</b> |
| Personality <sup>b</sup>         | 0.24 ± 0.46         | 0.51         | 0.62            |
| Sex <sup>c</sup>                 | -0.49 ± 0.50        | -0.97        | 0.35            |
| Trial number                     | <b>-0.11 ± 0.51</b> | <b>-2.09</b> | <b>&lt;0.05</b> |
| Previous experience <sup>d</sup> | -0.45 ± 0.28        | -1.59        | 0.11            |
| Inhibition score                 | -0.01 ± 0.07        | -1.20        | 0.84            |

*N* = 128, *df* = 104. Data used in this model are from the subset of 22 individuals tested for inhibition score. Random effect included bird identity (1.24; 95% confidence interval, CI = 1.09, 1.42) and the residual variance (0.51; 95% CI = 0.28, 0.95). Significant results (*P* < 0.05) are highlighted in bold.

<sup>a</sup> Low (reference level is high).

<sup>b</sup> Slow (reference level is fast).

<sup>c</sup> Male (reference level is female).

<sup>d</sup> Yes (reference level is no).

**Table A7**

Full model outputs from GLMM with factors affecting accuracy per trial

| Fixed effects                 | $\beta \pm SE$     | <i>t</i>    | <i>P</i>         |
|-------------------------------|--------------------|-------------|------------------|
| Intercept                     | 0.21 ± 0.23        | 0.88        | 0.38             |
| Interaction rate              | 0.00 ± 0.00        | 0.81        | 0.42             |
| Previous experience           | <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             |

*N* = 126, *df* = 99. Data used in this model are from the subset of 22 individuals tested for inhibition score. Random effect included bird identity (0.24; 95% confidence interval, CI = 0.15, 0.36) and the residual variance (0.19; 95% CI = 0.13, 0.30). Significant result (*P* < 0.05) is highlighted in bold.

<sup>a</sup> Slow (reference level is fast).

<sup>b</sup> Low (reference level high).

<sup>c</sup> Male (reference level female).

**Table A8**

Full model outputs from GLMM with factors affecting solving access points within trials

| Fixed effects                     | $\beta \pm SE$      | <i>z</i>     | <i>P</i>         |
|-----------------------------------|---------------------|--------------|------------------|
| Intercept                         | -4.45 ± 1.49        | -2.98        | <0.001           |
| Accuracy                          | <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             |
| Motivation group <sup>b</sup>     | <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             |

*N* = 126, *df* = 115. Data used in this model are from the subset of 22 individuals tested for inhibition score. Random effect included bird identity (<0.01; 95% confidence interval, CI = 0.0, 0.0) and the residual variance (<0.01; 95% CI = 0.0, 0.0). Significant results (*P* < 0.05) are highlighted in bold.

<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> Male (reference level is female).

**Table A9**

Full model outputs from GLM with factors affecting the number of different access points solved by an individual

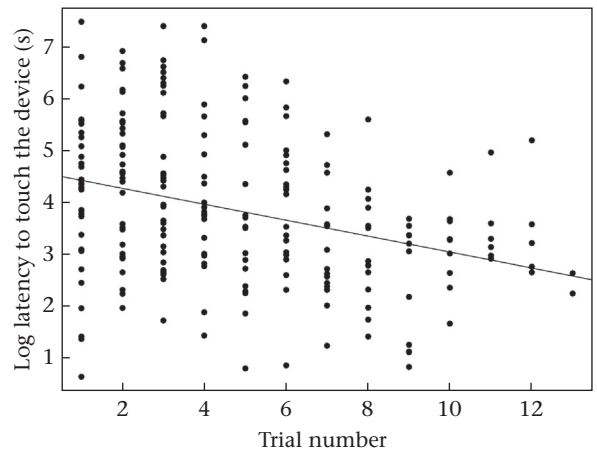
| Explanatory variables         | $\beta \pm SE$      | <i>z</i>     | <i>P</i>        |
|-------------------------------|---------------------|--------------|-----------------|
| Intercept                     | -0.67 ± 1.13        | -0.59        | 0.55            |
| Motivation group <sup>a</sup> | <b>-0.11 ± 0.53</b> | <b>-1.99</b> | <b>&lt;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            |

*N* = 22, *df* = 16. Data used in this model are from the subset of 22 individuals tested for inhibition score. Near-significant result is highlighted in bold.

<sup>a</sup> Low (reference level is high).

<sup>b</sup> Slow (reference level is fast).

<sup>c</sup> Male (reference level is female).



**Figure A1.** Latency to touch the device in relation to trial number. Data are log transformed for scale purposes.