Inhibitory control performance is repeatable over time and across contexts in a wild bird population

Gabrielle L. Davidson [a, b, *], Michael S. Reichert [a, c], Jenny R. Coomes [a],
Ipek G. Kulahci [a, d], Iván de la Hera [a, c], John L. Quinn [a, e, *]

[a] School of Biological, Earth and Environmental Sciences, University College Cork, Cork, Ireland
[b] Department of Psychology, University of Cambridge, Cambridge, U.K.
[c] Department of Integrative Biology, Oklahoma State University, Stillwater, OK, U.S.A.
[d] University of Notre Dame, Notre Dame, IN, U.S.A.
[e] Environmental Research Institute, University College Cork, Cork, Ireland

ARTICLE INFO

Inhibitory control is one of several cognitive mechanisms required for self-regulation, decision making and attention towards tasks. Inhibitory control is expected to influence behavioural plasticity in animals, for example in the context of foraging, social interaction or responses to sudden changes in the environment. One widely used inhibitory control assay is the 'detour task' where subjects must avoid impulsively touching transparent barriers positioned in front of food, and instead access the food by an alternative but known route. However, because the detour task has been reported to measure factors unrelated to inhibitory control, including motivation, previous experience and persistence, the task may be unreliable for making cross-species comparisons, estimating individual differences and linking performance with socioecological traits. To address these concerns, we designed a variant of the detour task for wild great tits, *Parus major*, and deployed it at the nesting site across two spring seasons. We compared task performance of the same individuals in the wild across 2 years, and with their performance in captivity when tested using the classical cylinder detour task during the nonbreeding season. Potential confounds of motivation, previous experience, body size, sex, age and personality did not significantly predict performance, and temporal and contextual repeatability were low but significant. These results support the hypothesis that our assays captured intrinsic differences in inhibitory control. Instead of dismissing detour tasks and 'throwing the baby out with the bathwater', we suggest confounds are likely system and experimental-design specific, and that assays for this potentially fundamental but largely overlooked source of behavioural plasticity in animal populations, should be validated and refined for each study system.

Inhibitory control is a form of self-regulation that allows an individual to supress predispositions in favour of more appropriate actions (Diamond, 2013). It is a domain-general cognitive ability and one of several executive functions required for decision making and attention towards tasks. Inhibitory control is increasingly being used to explain sociality and functional behaviour (reviewed in Kabadayi et al., 2018). This includes foraging flexibility when predation risk varies (Coomes et al., 2021), dietary breadth in individuals and species (MacLean et al., 2014; van Horik et al., 2018) and inhibiting previously employed social strategies depending on the identity of social group members (Amici et al., 2008; 2018; Reddy et al., 2015). Moreover, inhibitory control is likely to be responsive to selection given its links with brain size in primates and its heritability in humans (Friedman et al., 2008), dogs, *Canis lupus* (Gnanadesikan et al., 2020), fish (Lucon-Xiccato & Bertolucci, 2020) and birds (Langley et al., 2020). However, the inferential power of these studies is dependent on cognitive assays that accurately characterize inhibitory control and whether these assays are applicable to how animals behave in natural settings.

Recently, researchers have taken a variety of classical cognitive tasks to the field (e.g. Johnson-Ulrich et al., 2020; Morand-Ferron et al., 2011, 2015, 2016; Muth et al., 2018; Reichert et al., 2020; Sonnenberg et al., 2019; Toledo et al., 2020) where cognition can be

* Corresponding authors.
  E-mail addresses: gd339@cam.ac.uk (G. L. Davidson), j.quinn@ucc.ie (J. L. Quinn).

assayed under natural conditions with high ecological validity. Likewise, depending on the scope of the planned research and institutional policies, welfare concerns and administrative burdens associated with housing animals in captivity are reduced under natural settings. However, inhibitory control assays in the wild are scant and rely on experimenters in close proximity to habituated animals (Ashton et al., 2018; Shaw et al., 2015). Moreover, the extent to which extraneous variables contribute to, or confound, cognitive performance may differ for captive-bred, temporarily captive wild individuals and free-living animals (Morand-Ferron et al., 2016), but comparisons across such experimental contexts have rarely been made for any cognitive trait (see Benson-Amram et al., 2013; Cauchoix et al., 2017; Forss et al., 2015; McCune et al., 2019; Morand-Ferron et al., 2011; Mouchet & Digemanse, 2021 for some examples).

One way of measuring inhibitory control is the widely used 'detour' task. In this task, subjects must avoid and move around a transparent barrier to retrieve a food reward that is positioned directly in front of them, but behind the barrier (Diamond, 1990). A central premise of this task is that the visible reward generates a strong prepotent impulse to approach it directly. This impulse must be inhibited, and the subject must initially move in a direction away from the reward to detour around the barrier successfully. The classic detour task is dependent on an experimenter rebaiting the apparatus with a food reward, such that its scope for field-based assays has been restricted to wild, human-habituated New Zealand robins, *Petroica australis*, and Australian magpies, *Gymnorhina tibicen* (e.g. Ashton et al., 2018; Shaw, 2017). Another potential limitation for cognitive tasks in the field is that food rewards may be accessible to multiple individuals at once (but see Cauchard et al., 2013, 2017; Shaw et al., 2015). Consequently, individual performance is likely to be subject to social interference, for example through social learning or competition from other group members (Cole & Quinn, 2012; Morand-Ferron & Quinn, 2011; Reichert et al., 2020).

Attributing a cognitive capability to an individual is potentially challenging if the cognitive trait of interest is confounded by some other factor, such as motivation, personality and experience (Morand-Ferron et al., 2016). In the case of inhibitory control, experience with experimental contingencies (e.g. odour, Santacà et al., 2019; human training history, Fagnani et al., 2016, Jelbert et al., 2016), apparatuses (e.g. transparent plastic, Santacà et al., 2019) and motor routines (van Horik et al., 2020) can predict performance, as can motivational factors such as body condition as a proxy for energetic state (Shaw, 2017). Animal personality traits, for example consistent differences between individuals in how readily they engage with and explore their environment, have been shown to affect how animals perform on cognitive tasks generally (Dougherty & Guillette, 2018; Guillette et al., 2017; Sih & Del Giudice, 2012), including on tests of inhibitory control (Gomes et al., 2020; Lucon-Xiccato et al., 2020; Savaşçı et al., 2021), although these effects are not universal (Guillette et al., 2015; Vernouillet et al., 2018). The extent to which cognitive tasks that target inhibitory control tap into additional cognitive and noncognitive processes may also be dependent on the type of task used (e.g. Völter et al., 2018), the context (e.g. captivity versus the wild) and temporal aspects (e.g. season). Therefore, pinpointing confounds and correlates is a critical step towards accurately assaying inhibitory control for any given study system.

Examining the consistency of task performance has the potential to shed light on task validity. Typically, consistency is quantified by estimating repeatability, that is, the proportion of variance in a repeatedly measured trait that is explained by between-individual differences. Significant repeatability suggests consistent individual differences in behaviour, and although it sets the upper limit of heritability (Dohm, 2002), equally these individual differences could be caused by permanent environmental effects (Wilson, 2018). Of the few studies that have explicitly examined repeat performance using the detour task, evidence for consistency between individuals is mixed. Captive zebrafish, *Danio rerio*, were consistent in their performance on the same task (Lucon-Xiccato & Bertolucci, 2020), but captive song sparrows, *Melospiza melodia*, captive pheasants, *Phasianus colchicus*, and wild New Zealand robins were not (Shaw, 2017; Soha et al., 2019; van Horik et al., 2018). Across different tasks (i.e. different types of transparent apparatuses), captive zebrafish performance was correlated, and wild Australian magpies showed significant repeatability (Ashton et al., 2018). Repeatability may not necessarily be strong proof of task validity since the task could be consistently measuring the same confounds, especially when using the same task under the same conditions over time. For example, in a recent study on problem solving in great tits, *Parus major*, performance was repeatable but this repeatability was largely explained by experimentally manipulated motivational effects that were present in both repeat measures (Cooke et al., 2021). This is less likely to be a problem, however, when repeat measures are taken under very different conditions (contextual repeatability) that are unlikely to have shared environmental confounds, and especially when intrinsic confounds are also likely to differ (Nakagawa & Schielzeth, 2010). Finally, as discussed above, confounding variables can explain variation in single measures of cognitive performance, but can also act on the between- and/or within-individual variance component when dealing with multiple measures from the same individual. It follows that a further important test is to establish whether the consistent between-individual differences specifically are driven by these confounding effects (Cooke et al., 2021; Nakagawa & Schielzeth, 2010). Demonstrating temporal and contextual repeatability in performance on the detour task would provide compelling support for a common cognitive basis of these tasks, especially when controlling for potential confounds. Our aim here was to examine the extent to which variations in the detour task assay inhibitory control in our study system of wild great tits across time and contexts.

We designed a modified version of the detour task and presented it at the nestbox over two breeding seasons. The task involved minimal experimental disturbance or social interference from conspecifics, and no need for food rewards. Additionally, motivational factors related to brood age, size or satiety were not expected to confound our task performance (Cauchard et al., 2017). We also ran a classic version of the task by testing wild-caught great tits in captivity during one winter season. This approach allowed us to test for both contextual and temporal repeatability, and to examine statistically whether environmental or state variables caused between-individual differences in detour task performance, rather than the hypothesized cognitive mechanisms underpinning inhibitory control. These variables included experience, motivation, body size (because large birds may have more difficulty avoiding the barrier) and habitat as potential ecological confounds linked to differences in population density (O'Shea et al., 2018) and/or environmental variability (van Horik et al., 2019). Moreover, we also examined whether 'exploration behaviour in a novel environment', a commonly used assay of the fast–slow exploration personality axis (Dingemanse et al., 2002), explained performance on the captive detour task. If these variables did not explain task performance in general, and the repeatability of performance in particular, this would lend support for the detour task's utility as a robust measure of inhibitory control, where individual performance is not sensitive to bias from extraneous influences.

METHODS

Study System and Field Sites

The great tit is a model species in ecology and evolution field studies and has also been used for exploring the evolutionary ecology of cognitive variation. Great tits breed in nestboxes and can be fitted with passive integrative transponder (PIT) identification tags for remote detection at nestboxes using radio frequency identification (RFID) technology. They readily engage with experimental apparatuses in the wild and in captivity (Aplin et al., 2015; Cauchard et al., 2013; Cole et al., 2012; Morand-Ferron et al., 2011; Troisi et al., 2021), allowing for comparison of cognitive performance across different settings. Our study took place across 10 distinct woodland sites in the Bandon Valley area, Co. Cork, Ireland. Five sites were mixed deciduous and five were conifer plantations (Table A1). Nestboxes, hung approximately 1.5 m from the ground, were distributed across these sites at a density of two nestboxes per ha. Nestboxes were monitored for breeding data (O'Shea et al., 2018; lay dates, incubation period, hatching date, clutch size and number of fledglings). Adults were trapped at the nestbox between day 10 and day 13 posthatching to measure biometrics (including weight, wing and tarsus length), and to tag individuals with a PIT tag and a coloured ring with a unique alphanumerical code for identification. Chicks were ringed and weighed on day 15 posthatching. Brood size was recorded on experimental dates.

Figure 1. The phases of the detour task (a, b) in the wild and (c, d) in captivity. (a, c) Training phases with an opaque barrier and (b, d) test phases with a transparent barrier. Perches were positioned below the nestbox entrance hole, and in front of the cylinders. For the habituation phase in the wild, the opaque barrier was inverted 180° so it was positioned above the box. For the habituation phase in captivity, the reward was placed visibly on the side of the tube.

Experimental Procedures: Detour Task in the Wild

To measure individuals' detour task performance in the wild, we designed a detour task for the nestbox requiring birds to avoid initially opaque, and subsequently transparent, flexible barriers to gain access to their nest hole. To do so, birds were required to alter their normal flight path (i.e. level with the nest hole), such that they could enter from the gap under the barrier (Fig. 1). The experiment consisted of three phases: (1) habituation to the opaque experimental apparatuses, (2) training to go under an opaque barrier and (3) testing with a transparent but otherwise similar barrier (Fig. 2). In both years, a Perspex cover (20 cm x 20 cm) was positioned horizontally on top of the box during the training and test phases, to provide birds with experience with transparent plastic (see Isaksson et al., 2018; van Horik et al., 2018) and to act as a cover for the transparent barrier in case of rain (although most experiments occurred when there was no rain). Experiments took place between 25 May and 25 June 2017 (hereafter year 1), and between 26 May and 15 June 2018 (hereafter year 2).

At the start of an experimental session, that is, when the experimental apparatus was placed on the nestbox for each phase, the standard nestbox front was replaced with an identical-looking front with integrated photodiode sensors and RFID technology that logged visits from birds with or without PIT tags (RFID logger for Schwegler 1B from Dominic Goodwin, NatureCounters, www.naturecounters.com). A Panasonic HC camera was mounted on a 50 cm high tripod, positioned approximately 10 m from the nest. Both the nest front and the camera were synchronized and used to quantify the number of times each individual bird landed on/entered the nest hole, and to identify the parents by colour ring and/or plumage characteristics. Each experimental session was presented at the nestbox for approximately 1 h each day over consecutive days between 0700 and 1300 hours. Three boxes received a test experiment between 1300 and 1600 hours (one box in year 1, two boxes in year 2). RFID and photodiode data were reviewed on the same day to determine whether birds passed the habituation and training phases (Fig. 2, see Appendix for criterion details). In year 1 the video footage confirmed the accuracy of the RFID data at the habituation and training phase, so that in year 2 only RFID data were required. Video footage was reviewed to quantify performance at the test phase in both years.

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

If neither parent passed a phase after 2 days, the experiment was abandoned for that nest. If only one parent passed the phase after 2 days, then the experiment was advanced for the participating parent. Therefore, the experiment took three to six experimental sessions (i.e. days) to complete in year 1 (mean =3.7
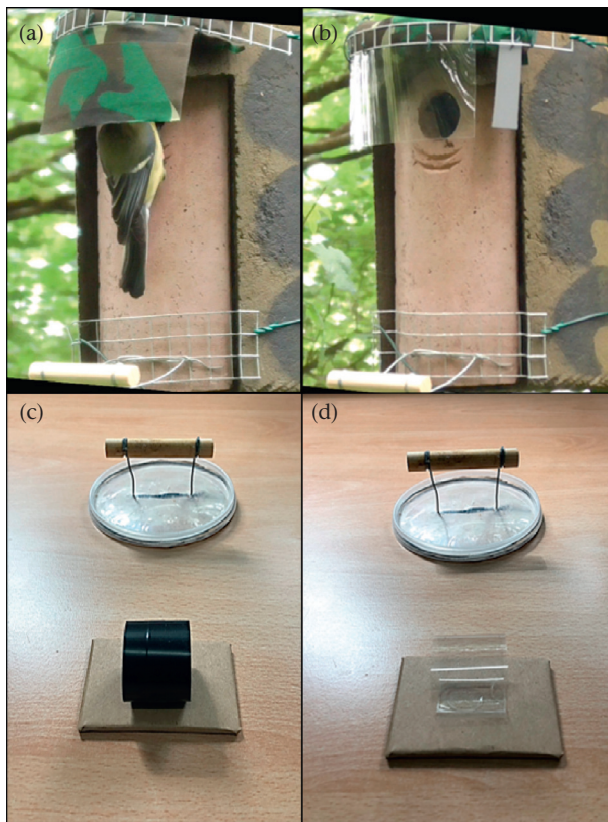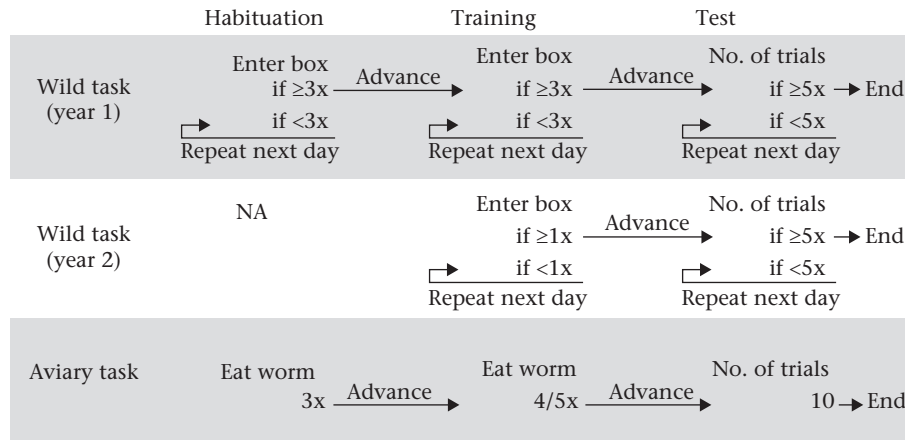
Figure 2. Schematic of the experimental phases (Habituation, Training, Test) for the wild tasks (year 1 and year 2) and the aviary task. Birds advanced to the next phase if they met the criterion, indicated by either entering the box or eating the worm the required number of times (indicated by 'x'). If they did not meet the criterion, the phase was repeated the following day. The criterion for the wild task in year 2 was reduced to minimize disturbance at the box. The experiment ended when birds performed the required number of trials in the test phase. Further details are provided in the Appendix.

days ±0.13 SE). None of the birds abandoned the nest during or after the experiments. Experiments started at day 10 posthatching, except for 12 boxes in year 1 that started at day 5–7 due to logistical constraints. Experiments did not occur on the trapping day. In year 2 the experimental protocol was refined to reduce the number of experimental trials by eliminating the need for a habituation phase and reducing the criterion for the training phase. Instead of the 1 h habituation phase described above, a dummy apparatus was placed on the box permanently from day 8 posthatching until the start of the training phase. This consisted of wire mesh around the top of the box attached to a plastic rectangle covered with camouflage tape (0.2 x 6 cm and 5 cm high), and a wooden perch below the nest hole. The criterion for passing the training phase was reduced to one successful approach, instead of three, as our year 1 data showed that the number of training trials did not influence performance during the test phase (see also Results). The experiment took 2–4 days to complete in year 2 (mean = 2.4 days ± 0.12 SE).

Aviary Housing

Birds were caught from nine of the 10 sites described above between January and early March 2018, four of which were mixed deciduous woodlands and five were conifer plantations. Birds were transported from the field sites to the aviary in cloth bags within 2 h of being caught and brought into captivity for approximately 3 weeks before being released at the site of capture. Housing details are described in Coomes et al. (2021). Birds were not deprived of food prior to experiments, and wax worms, *Galleria mellonella* larvae, were only provided as experimental rewards.

Experimental Procedures: Detour Task in Captivity

The captive detour task consisted of the same three experimental phases described above (Fig. 2). We piloted different sizes of tubes on a cohort of birds not included in the main analyses and chose a 3.5 cm diameter x 3 cm wide cylinder tube so that the complexity of the task did not cause ceiling or floor effects in performance caused by the difficulty in detouring around the barrier (Farrar et al., 2020; Völter et al., 2018). A 5 cm high perch was positioned 15 cm in front of the cylinder to standardize the approach direction for each trial (Fig. 1). The habituation phase was presented the day after birds arrived in the aviary. Birds had to retrieve a wax worm placed at the edge of an opaque plastic

cylinder three times before they received the training phase. Wax worms were euthanized by head compression, so they did not move during the trial. Depending on the bird's progress, the training phase occurred either on the same day or the following day.

In the training phase, birds had to retrieve a wax worm placed in the centre of the same plastic cylinder, without touching the exterior of the cylinder, on four of five consecutive attempts to retrieve the worm (Boogert et al., 2011). The test phase was always performed the day after a bird passed the training phase. During this phase birds received 10 trials with a transparent cylinder of the same dimensions as the one in the habituation and training phases. To ensure birds did not become sated, half a wax worm was used as a reward during the test trials. A trial was defined as a bird approaching the cylinder and making contact with the barrier (scored as a fail) or retrieving the worm from the side of the tube without making contact with the barrier (scored as a success). The trial ended when the bird retrieved the worm, or flew away from the apparatus, at which point the tube was removed from the testing enclosure, rebaited by the experimenter, and placed back in the enclosure. This procedure was designed such that each approach was measured as a success/fail, as opposed to the number of pecks until success, the latter of which may be guided by individual persistence (van Horik et al., 2018). Allowing birds to consume the worm at each trial, whether they failed or succeeded, controlled for reward history that may have influenced reinforcement and/or motivation through hunger. All birds ate 8–10 worms at the test phase, except two birds, which ate four and seven worms. We recorded the time it took each bird ($N=35$) to complete 10 trials as there was no limit to how long birds had to approach the cylinder for each trial. As for the wild task, we measured the captive detour task performance as the proportion of successful trials out of total trials, such that high values indicated putatively high inhibitory control.

Experimental Procedures: Exploration Behaviour

The morning following the birds' arrival to the aviary, we performed an 'exploration in a novel environment' assay, henceforth referred to as exploration behaviour (see also Coomes et al., 2021; adapted from; Dingemanse et al., 2002). An access hatch at the back of the bird's cage that led to a larger room (4.60 x 3.10 m and 2.65 m high) was opened. The light in the home cage was turned off and

birds were free to enter the room. Once the bird entered the room, the number of hops and flights, within and between trees, was recorded from the adjacent corridor through one-way glass. 'Trees' were made of a wooden upright support and two thick dowels running at right angles to each other (see also Coomes et al., 2021). The trial was complete after 2 min, at which point the birds were returned to their home cage. Exploration behaviour was recorded as the sum of the number of hops and flights, and has been shown to be repeatable in our population (O'Shea et al., 2017).

Statistical Analysis: General Comments

All models, unless otherwise specified, were run as generalized linear (mixed) models (GLMMs) in lme4 (Evans, 2016) in the R statistical software interface (R Core Team, 2017). *P* values were generated using lmerTest (Kuznetsova et al., 2017). Plots were generated using ggplot (Wilkinson, 2011). We used the dredge function from the MuMIn package (Barton 2019) and an information-theoretic approach in combination with model averaging (Grueber et al., 2011). We generated models from a global model from our GLMMs and retained models with an Akaike's information criterion corrected for small sample sizes (AICc) within 2 units of the top model. We report the conditional averaged weighted parameter estimates across the retained models. All continuous variables were scaled. We used the vif function in the usdm package (Naimi et al., 2014) to test for collinearity between fixed factors. All variables had a variance inflation factor less than 2.5 and were not considered to show multicollinearity. Our R code is included as Supplementary material.

Statistical Analysis

Detour task in the wild
The number of trials undertaken by each bird during the 1 h test phase varied (mean ± SE trials = 9.6 ± 0.57 SE, range 1–26). Our analyses included a maximum of the first 10 trials as this was consistent with the number of trials in the captive task and existing literature (MacLean et al., 2014). We also confirmed that the number of trials used to calculate overall performance did not bias wild detour task performance, if, for example, birds with more trials had higher scores if they learned over successive trials to avoid the barrier (see Results).

Initially we examined what fixed effects had a potentially confounding influence on detour task performance in the wild. Lack of any strong effects would lend support for the wild detour task being a reliable test of inhibitory control. It is also important to identify which fixed effects could be driving between-individual variation in detour task performance (see repeatability below). We modelled wild detour task performance in a GLMM with a binomial distribution and logit link function, with the number of successes as the numerator and the total trials as the denominator (in R, using cbind, the response variable is entered as two variables, number of successes and number of fails). Our global model included the following fixed effects as potential sources of variation in performance: the number of training trials because the motor action of flying under a barrier could carry-over to the test phase; wing length as an index of body size that could influence the ability to pass underneath the barrier; year, lay date and brood size as these could be sources of motivation that may influence parental impulses to feed their chicks; and sex, which has been reported to predict inhibitory control performance in a stop-signal task (Lacreuse et al., 2016). Continuous variables were scaled and mean-centred to zero. Site, nest and bird identity (ID) were included as nested random terms. Owing to convergence issues associated with overfitting the model with categorical variables, we did not include

habitat (conifer versus deciduous) or age in our global model, although visual inspection of these variables and reduced models in which these variables were included suggest they had no effect on performance (Fig. A1).

Detour task in captivity
We modelled captive detour task performance in a GLMM with a binomial distribution and logit link function. Our global model included reward history (i.e. number of worms eaten), motivation (i.e. time to complete all 10 test trials), habitat, personality, sex and age as fixed factors, and site of capture as a random effect. Habitat was included as a potential ecological confound linked to differences in population density (O'Shea et al., 2018) and/or environmental variability (van Horik et al., 2019). Exploration behaviour was included as a fixed factor as personality may influence how birds engage with the task, or form part of so-called cognitive 'styles' (Sih & Del Giudice, 2012). Sex and age were also included as sources of variation in task performance (Lacreuse et al., 2016; Macdonald et al., 2014).

Repeatability
We investigated whether individuals showed consistent differences in the wild detour task performance across years (temporal repeatability; but not for the captive task for which we had no repeats). Significant temporal repeatability in performance would suggest that the task measured an intrinsic trait, indicating a permanent environment effect and/or heritability (e.g. Quinn et al., 2009). The wild detour task data set included repeat measures ($N = 16$ observations, eight individuals), as well as single measures ($N = 68$) to increase power (Martin et al., 2011). We ran a GLMM as described above, with year as a fixed effect and bird ID as a random effect, and compared this model with another that excluded bird ID using the anova function in R. The repeatability estimate was calculated from the variance components and the residual variance as $1/p(1-p)$, where p is the expected probability of success calculated as the mean wild detour task performance in the data set (Nakagawa et al., 2017). The 2.5% and 97.5% confidence intervals (CI) were calculated with the function confint() using the bootstrap argument with 1000 simulations. Sigma (residual deviation) was estimated to be 1. We also tested whether the inclusion of fixed effects resulted in any change in the repeatability estimate (adjusted repeatability). If the repeatability of wild detour task performance remained significant after inclusion of these effects, this would point to consistent individual differences being explained by inhibitory control.

We also estimated contextual repeatability between tasks, significant levels of which would suggest that performance on these tasks could be attributed to a common factor, supporting the hypothesis that these tasks reflect, at least in part, inhibitory control where the prepotent impulse to go directly towards a positive stimulus (either a food reward or a begging offspring) must be inhibited. We ran an additional two GLMMs (with and without bird ID as a random effect) using a data set that included repeat measures ($N = 21$ observations, 10 individuals, one of whom was measured both in year 1 and in year 2 of the wild task), and singletons ($N=98$). Differences in results were negligible if we excluded year 2 performance from the bird for which we had repeated measures in years 1 and 2 (see Appendix). We included task (wild versus captivity) as a fixed effect, and bird ID and site as random effects. We then repeated these analyses to control for fixed effects that were common between tasks and had been retained in the model selection for temporal repeatability, to ensure that any significant repeatability was not driven solely by common factors unrelated to inhibitory control.

Ethical Note

This study was conducted under licences from the Health Products Regulatory Authority (AE19130_P017), The National Parks and Wildlife Services (C11/2017, 004/2017, C02/2018 and 001/2018) and the British Trust for Ornithology. The research project received ethical approval from the Animal Welfare Body at University College Cork, and was in accordance with the ASAB/ABS Guidelines for the Treatment of Animals in Behavioural Research and Teaching. We observed no nest desertion as a result of our experiment.

RESULTS

Participation

In year 1, 19 males and 24 females from 29 nests participated in the wild detour task. The experiment was attempted at three additional nests, but these were excluded because neither parent passed the habituation or training phases. In year 2, 21 females and 20 males from 22 nests participated, where at least one parent from all nests reached and participated in the test phase. One bird did not participate in the captive detour task as it would not approach or consume the worm next to the apparatus.

We found that the total number of test trials did not correlate with overall performance (Kendall's tau test: $z = -0.23$, tau $= -0.02$, $P = 0.79$, $N = 84$, Fig. A2a). Moreover, for birds that completed at least 10 trials, their overall performance calculated from the first five trials was highly correlated with their overall performance calculated from the first 10 trials ($z = 9.42$, tau $= 0.86$, $P < 0.001$, $N = 69$, Fig. A2b). Eight individuals completed fewer than five trials and were included in the analysis (three birds completed one trial, one bird completed two trials, one bird completed three trials and three birds completed four trials).

Detour Task in the Wild

The number of instances that individuals successfully went under the opaque barrier during the training phase varied across individuals (mean $= 7.11 \pm 0.59$ SE, range 1−37). In year 1, during the training phase there were eight instances from seven individuals when birds made contact with the opaque barrier, whereas during the test phase there were 130 instances from 36 individuals when birds made contact with the transparent barrier (mean $= 3.02 \pm 0.45$ SE), confirming that the transparent barrier elicited the prepotent response of flying straight to the nest hole (see van Horik et al., 2018). Year, lay date, sex and wing length were retained in the top models, but there was no evidence that any influenced detour task performance (Fig. 3, Table 1), although males performed marginally worse than females, and there was a tendency for performance to decline with lay date. The number of training trials, brood size and year were not retained in any of the top models and did not predict task performance (Fig. 3, global model test statistics are provided in Table A2).

Detour Task in Captivity

The mean captive detour task performance was 0.41±0.04 SE, substantially lower than that observed in the wild task. The time it took birds to complete the test phase, the number of worms eaten, sex and exploration behaviour were the retained fixed terms in the top models, but evidence that any of these variables had an effect was weak because none were statistically significant (Fig. A3). Age (Fig. A3c) and habitat (Fig. A3f) were not retained in any of the top models and did not predict task performance (global model test statistics are provided as Table A2).

Repeatability

The temporal repeatability of wild detour task performance across years was low but significant ($R = 0.19$, $P <0.001$, CI $= 0.06$, 0.22), and remained significant when controlling for sex, lay date and wing length (adjusted $R = 0.15$, $P < 0.001$, CI $= 0.001$, 0.18; Fig. 4a, c). Contextual repeatability in performance across tasks was low but significant ($R = 0.18$, $P < 0.001$, CI $= 0.09$, 0.24), and remained significant when controlling for sex (adjusted $R = 0.17$, $P < 0.001$, CI $= 0.001$, 0.22; Fig. 4b, d). Task type (wild, captive), which was included in the contextual repeatability models, was significant: performance in the wild was significantly higher than performance in captivity ($z = 4.34$, $P < 0.001$).

DISCUSSION

We have shown that, in wild great tits, the detour task is repeatable across years and testing environments (wild versus captivity). Furthermore, we controlled for a range of possible confounding variables, and they did not explain the repeatability of performance. Although our sample size was small, these results suggest an underlying trait that is common across tasks in our system, which we interpret as the inhibition of a prepotent response/habitual behaviour. To our knowledge, two other studies have investigated repeatability of inhibitory control in wild birds (Ashton et al., 2018; Shaw, 2017), only one of which also reported significant repeatability (Ashton et al., 2018). Moreover, our findings contrast with recent reports that the detour task is correlated with cognitive and noncognitive traits unrelated to inhibitory control (Shaw, 2017; van Horik et al., 2018; van Horik et al., 2020). By synthesizing our findings across both wild and captive versions of the detour task, we discuss the extent to which we can attribute task performance to inhibitory control.

Performance on the wild detour task was repeatable across time, and between wild and captive versions of the task. Although repeatability estimates were low, these results point to an underlying, inherent trait that was consistently measured despite differences in year, season, testing location and apparatus type. Repeatability sets the upper limit for heritability but does not preclude permanent environment effects driving some or even all the intrinsic differences observed. Nevertheless, genetic pedigree studies have shown that detour task performance is heritable in birds (Langley et al., 2020), and that executive functioning is the most heritable psychological trait in humans (Friedman et al., 2008). While we acknowledge that the relatively small sample size of within-individual measurements may render parameter estimation less reliable (Nakagawa & Schielzeth, 2010), our repeatability estimates were consistent with previous reports in animal cognition. A meta-analysis of repeatability estimates across a range of different cognitive tasks including inhibitory control, problem solving, discrimination and reversal learning, memory, physical and spatial cognition found low to moderate R values for temporal repeatability (0.15 and 0.28) and contextual repeatability (0.20−0.28; Cauchoix et al., 2018). The two detour task studies in the meta-analysis had very opposing results: no repeatability in New Zealand robins ($R = 0.002$; Shaw, 2017) but high repeatability in Australian magpies ($R = 0.80$; Ashton et al., 2018). This suggests that repeatability of the detour task across species and populations may be considerably variable if temporary environmental effects vary across space and time. The particularly high repeatability reported for the Australian magpies, for example, is likely due to the repeats being taken just 2 weeks apart (in comparison to the 12 months in this study) which is likely to inflate differences between individuals due to transient factors (Bell et al., 2009; Cole et al., 2011). In addition, cognitive tasks that quantify performance as
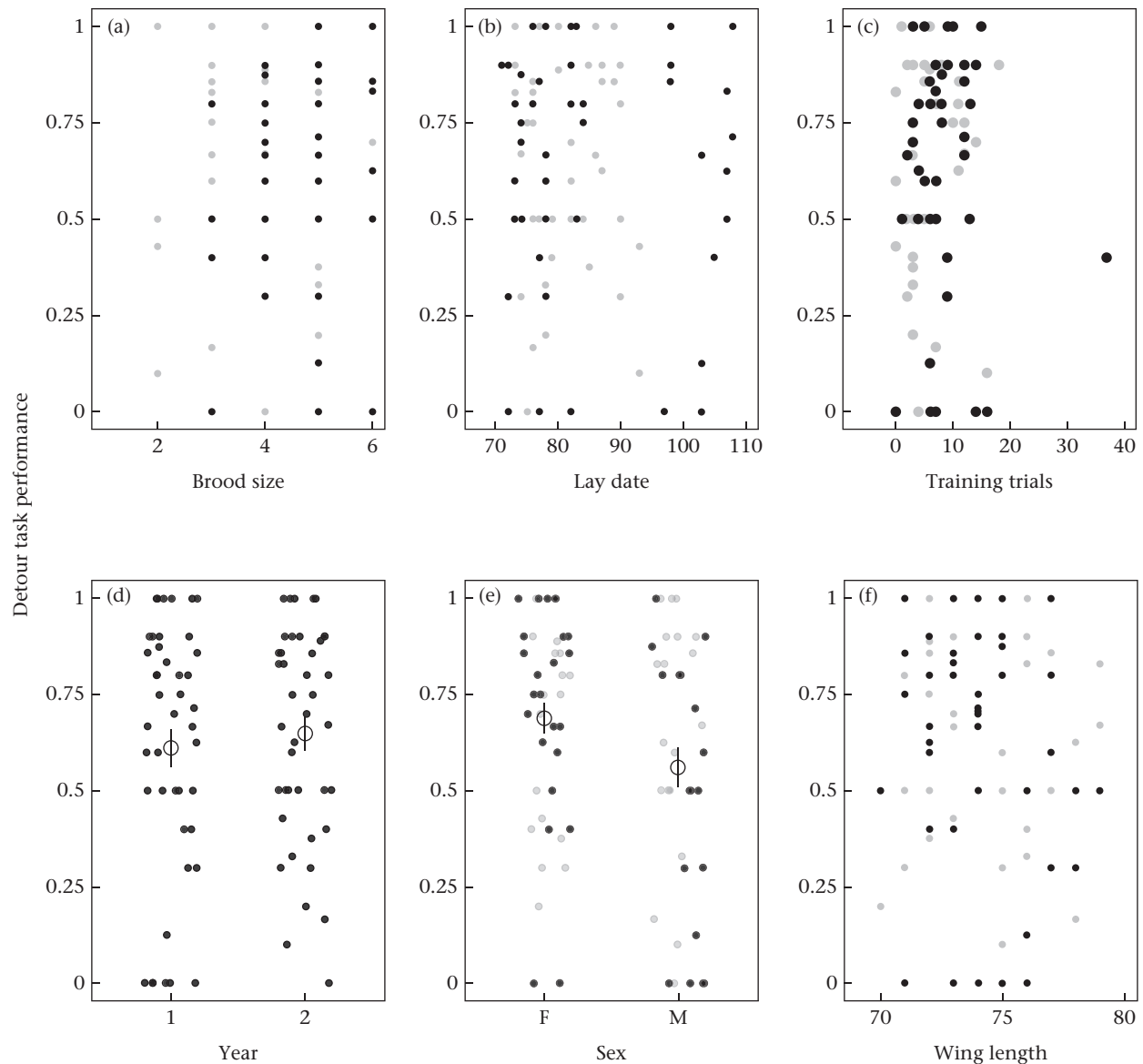
Figure 3. Detour performance (the proportion of successful trials out of total trials, where higher values indicate better performance) in the wild in relation to (a) brood size, (b) lay date (days since 1 March), (c) total number of training trials, (d) year, (e) sex (F=female, M=male) and (f) wing length (mm). Points represent individuals: black points were taken in year 1 and grey in year 2. Data points for (c), (d) and (f) are jittered along the horizontal axis to reduce overlap. Open circle and line represent mean ± SE across both years. See Table 1 for model results.

Table 1
Binomial GLMMs of detour task performance

	Estimate ± SE	z	P
Wild detour task			
Intercept	0.96±0.38	2.54	0.01
Lay date	−0.34±0.20	−1.65	0.10
Sex (Female)	−0.867±0.49	−1.73	0.08
Wing length	0.35±0.27	1.27	0.20
Captive detour task			
Intercept	−0.32±0.20	−1.57	0.11
Time to complete test	0.14±0.12	1.23	0.25
Sex (Female)	−0.24±0.24	−0.96	0.34
Exploration behaviour	0.09±0.13	0.66	0.51
Total rewards eaten	0.19±0.13	1.36	0.17

Performance was measured as the proportion of successful test trials for the wild detour task (N=84) and the captive detour task (N=35). The values shown are the conditional average of the top models within two AICc of the best model. All continuous variables were scaled. Reference categories for binary variables are in parentheses.

binary success/fail achieve higher R estimates than those that tally the number of trials to reach criterion (Cauchoix et al., 2018), and therefore, without an additional measure of performance to compare our success/fail measurement, repeatability estimates here may be an overestimate. Differences in performance and effect sizes are to be expected across detour task studies, and comparative cognition studies generally (Farrar et al., 2020). Our finding that the same individuals performed better on the wild than the captive task highlights how different variants of the detour task are likely to affect how well individuals perform, which makes comparisons difficult.

Experience (Fagnani et al., 2016; Santacà et al., 2019; van Horik et al., 2019, 2020), motivation (Shaw, 2017) and persistence (van Horik et al., 2018) have been proposed to confound individual differences in inhibitory control. Performance on the wild task was higher than performance on the captive task, perhaps because birds also experienced the transparent barrier as they exited the box, not
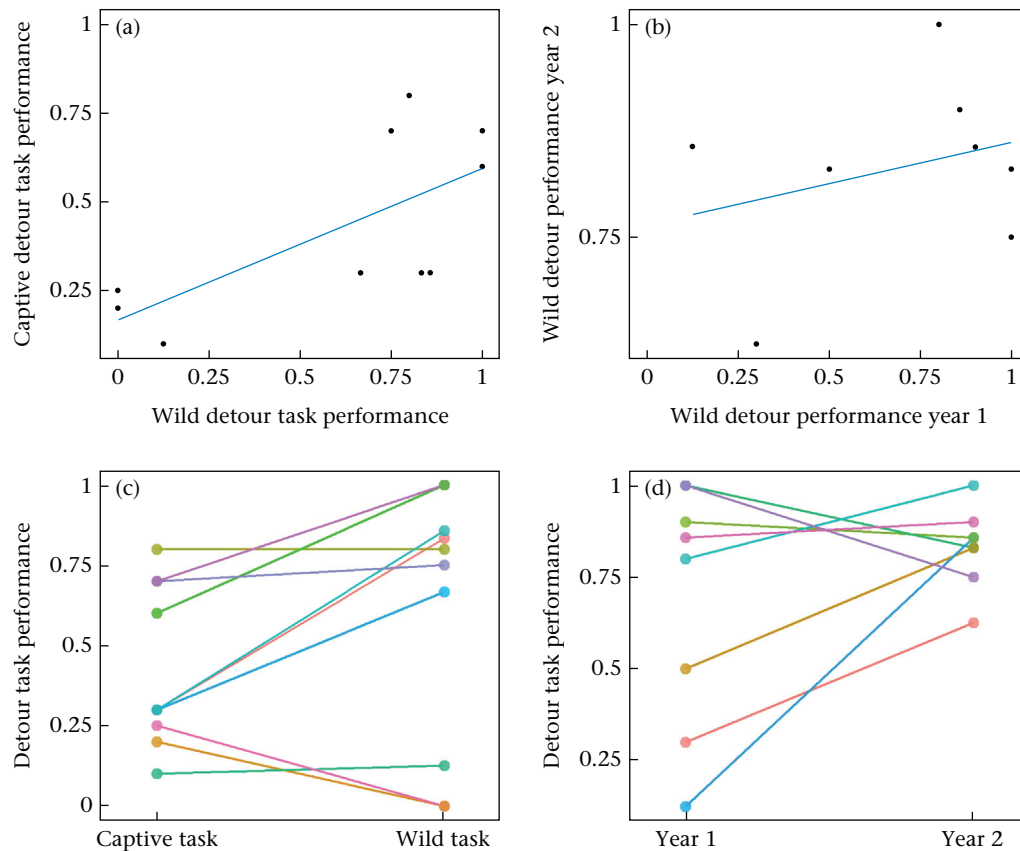
Figure 4. The relationship between individual detour task performance (the proportion of successful trials out of total trials, where higher values indicate better performance). (a) The wild and captive detour tasks (*N*=10 individuals). (b) Year 1 and year 2 wild detour tasks (*N*=8 individuals). Reaction norms for (c) contextual repeatability across captive and wild detour tasks and (d) temporal repeatability across years with the wild detour task. Circles represent individual performance; a regression line is fitted in (a) and (b) and lines connect repeated measures within individuals in (c) and (d). Lines and circles are in different colours per individual.

just when they approached it. Motivational factors could also explain differences in performance between the wild and captive tasks. The use of food rewards may contribute to motivational effects attributed to between-individual differences in hunger state, body condition and/or food preferences (Cooke et al., 2021; Shaw, 2017). Placing the detour task in front of the nestbox access hole allowed us to measure performance independently of food rewards and we expected that all individuals would be highly motivated to participate in and complete the inhibitory control tasks as quickly as possible. Nevertheless, motivation could vary depending on the reproductive value and the viability of each parent's offspring. However, we found no evidence for this because performance was unaffected by brood size and lay date. Although we did not control for chick age and satiety in our own study due to model over-paramaterization, these variables were not shown to influence great tit problem solving performance at the nestbox (Cauchard et al., 2017). Moreover, performance was repeatable despite different reward incentives between wild and captive tasks. Approach latencies, or time to complete a task, have been interpreted as proxies for motivation to obtain a reward, but we found no effect of these variables on performance in our captive task, nor did a similar study with common pheasants (van Horik et al., 2020). Body condition may reflect an individual's energetic state, and has been linked with performance in the detour task in wild New Zealand robins (Shaw, 2017). We did not test for such an effect in the current study as we did not have an accurate measure of body weight at the time of the experiments. Weights were only taken when birds were handled (i.e. at capture), and can fluctuate as much as 5% for great tits in captivity (G Davidson, personal

observation), and while parents are provisioning chicks. We controlled for many potential proxies of motivation and found that they did not affect performance, but it remains possible that there are intrinsic individual differences that we did not account for (Morand-Ferron et al., 2016).

Between-individual differences in animal personality may influence task performance if personality is intrinsically linked to so-called cognitive 'styles' (Sih & Del Giudice, 2012). For example, behaviours commonly attributed to the reactive–proactive animal personality axis include slow-exploring, environmentally sensitive individuals at one extreme and fast-exploring, routine-forming individuals on the other (Réale et al., 2007). These definitions have many parallels with definitions associated with human inhibition and impulsivity, including an impulsive behaviour with no forethought of consequences (Moeller et al., 2001). We found no evidence that exploratory behaviour, a common behaviour associated with the fast–slow exploration personality axis, was associated with performance on the detour task in captivity, thus excluding this as a dominant influence on cognitive performance. This is consistent with reports in black-capped chickadees, *Poecile atricapillus* (Guillette et al., 2017) and domestic dogs (Bray et al., 2015), but not common waxbills, *Estrilda astrild* (Gomes et al., 2020). We note that our measure of personality is an index of the fast–slow personality axis (Bell, 2007), and it may be that specific facets of this axis, for example responsiveness, the quality of exploration (i.e. information gathered) and neophobia, need to be measured in isolation to detect links with inhibitory control. Similarly, other kinds of personality axes, or behavioural variation in general, could play a role. Equally, we also note that our measure of inhibitory

control likely only captures one facet of self-regulation but there are many others (e.g. delayed gratification; e.g. Mischel et al., 1989) that themselves may be controlled by distinct but related forms of inhibitory control we did not measure here. Despite the challenges in teasing apart different elements of inhibitory control, and cognition generally, much of the psychology literature suggests that different measures of inhibitory control are a component of a wider latent cognitive variable, such as general inhibition, executive functioning and general intelligence (Anderson & Weaver, 2009; Aron et al., 2004; Bari & Robbins, 2013), which has also been demonstrated in wild systems (Shaw et al., 2015; Ashton et al., 2018).

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

Author Contributions

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

Acknowledgments

Supplementary Material

Supplementary material associated with this article is available, in the online version, at https://doi.org/10.1016/j.anbehav.2022.02.007.

References

Amici, F., Aureli, F., & Call, J. (2008). Fission-fusion dynamics, behavioral flexibility, and inhibitory control in primates. *Current Biology, 18*, 1415–1419. https://doi.org/10.1016/j.cub.2013.06.020

Amici, F., Call, J., Watzek, J., Brosnan, S., & Aureli, F. (2018). Social inhibition and behavioural flexibility when the context changes: A comparison across six primate species. *Scientific Reports, 8*(1), 3067. https://doi.org/10.1038/s41598-018-21496-6

Anderson, M. C., & Weaver, C. (2009). Inhibitory control over action and memory. In *Encyclopedia of Neuroscience* (pp. 153–163). https://doi.org/10.1016/B978-008045046-9.00421-6

Aplin, L. M., Farine, D. R., Morand-Ferron, J., Cockburn, A., Thornton, A., & Sheldon, B. C. (2015). Experimentally induced innovations lead to persistent culture via conformity in wild birds. *Nature, 518*(7540), 538–541. https://doi.org/10.1038/nature13998

Aron, A. R., Robbins, T. W., & Poldrack, R. A. (2004). Inhibition and the right inferior frontal cortex. *Trends in Cognitive Sciences, 8*(4), 170–177. https://doi.org/10.1016/j.tics.2004.02.010

Ashton, B. J., Ridley, A. R., Edwards, E. K., & Thornton, A. (2018). Cognitive performance is linked to group size and affects fitness in Australian magpies. *Nature, 554*(7692), 364–367. https://doi.org/10.1038/nature25503

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

Bell, A. M. (2007). Future directions in behavioural syndromes research. *Proceedings of the Royal Society B: Biological Sciences, 274*(1611), 755–761. https://doi.org/10.1098/rspb.2006.0199

Bell, A. M., Hankison, S. J., & Laskowski, K. L. (2009). The repeatability of behaviour: A meta-analysis. *Animal Behaviour, 77*(4), 771–783. https://doi.org/10.1016/j.anbehav.2008.12.022

Benson-Amram, S., Weldele, M. L., & Holekamp, K. E. (2013). A comparison of innovative problem-solving abilities between wild and captive spotted hyaenas, *Crocuta crocuta. Animal Behaviour, 85*(2), 349–356. https://doi.org/10.1016/j.anbehav.2012.11.003

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

Bray, E. E., MacLean, E. L., & Hare, B. A. (2015). Increasing arousal enhances inhibitory control in calm but not excitable dogs. *Animal Cognition, 18*(6), 1317–1329. https://doi.org/10.1007/s10071-015-0901-1

Cauchard, L., Angers, B., Boogert, N. J., Lenarth, M., Bize, P., & Doligez, B. (2017). An experimental test of a causal link between problem-solving performance and reproductive success in wild great tits. *Frontiers in Ecology and Evolution, 5*, 107. https://doi.org/10.3389/fevo.2017.00107

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., Bize, P., Boussard, A., Buechel, S. D., Cabirol, A., Cauchard, L., Claidière, N., Dalesman, S., Devaud, J. M., Didic, M., Doligez, B., Fagot, J., Fichtel, C., Henke-Von Der Malsburg, J., … Morand-Ferron, J. (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

Cauchoix, M., Hermer, E., Chaine, A. S., & Morand-Ferron, J. (2017). Cognition in the field: Comparison of reversal learning performance in captive and wild passerines. *Scientific Reports, 7*(1). https://doi.org/10.1038/s41598-017-13179-5

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. *Proceedings of the Royal Society B: Biological Sciences, 279*(1731), 1168–1175. https://doi.org/10.1098/rspb.2011.1539

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

Coomes, J. R., Davidson, G. L., Reichert, M. S., Kulahci, I. G., Troisi, C. A., & Quinn, J. L. (2021). Inhibitory control, exploration behaviour and manipulated ecological context are associated with foraging flexibility in the great tit. *Journal of Animal Ecology, 91*, 320–333. https://doi.org/10.1111/1365-2656.13600

Diamond, A. (1990). Developmental time course in human infants and infant monkeys, and the neural bases of, inhibitory control in reaching. *Annals of the New York Academy of Sciences, 608*(1), 637–676. https://doi.org/10.1111/j.1749-6632.1990.tb48913.x

Diamond, A. (2013). Executive functions. *Annual Review of Psychology, 64*(135), 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

Dingemanse, 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

Dohm, M. R. (2002). Repeatability estimates do not always set an upper limit to heritability. *Functional Ecology, 16*(2), 273–280. https://doi.org/10.1046/j.1365-2435.2002.00621.x

Dougherty, L. R., & Guillette, L. M. (2018). Linking personality and cognition: a meta-analysis. *Philosophical Transactions of the Royal Society B, 373*, Article 20170282. https://doi.org/10.1098/rstb.2017.0282

Evans, J. A. (2016). lme4: Linear mixed-effects models using eigen and S4. R package version 1.1-7. *Functional Ecology, 30*(7), 1–11. https://doi.org/10.5061/dryad.dn433

Fagnani, J., Barrera, G., Carballo, F., & Bentosela, M. (2016). Is previous experience important for inhibitory control? A comparison between shelter and pet dogs in A-not-B and cylinder tasks. *Animal Cognition, 19*(6), 1165–1172. https://doi.org/10.1007/s10071-016-1024-z

Farrar, B. G., Voudouris, K., & Clayton, N. S. (2020). Replications, comparisons, sampling and the problem of representativeness in animal behavior and cognition research. *Psyarxiv, 8*, 273–295. https://psyarxiv.com/2vt4k/.

Forss, S. I. F., Schuppli, C., Haiden, D., Zweifel, N., & van Schaik, C. P. (2015). Contrasting responses to novelty by wild and captive orangutans. *American Journal of Primatology, 77*(10), 1109–1121. https://doi.org/10.1002/ajp.22445

Friedman, N. P., Miyake, A., Young, S. E., DeFries, J. C., Corley, R. P., & Hewitt, J. K. (2008). Individual differences in executive functions are almost entirely genetic in origin. *Journal of Experimental Psychology: General, 137*(2), 201–225. https://doi.org/10.1037/0096-3445.137.2.201

Gnanadesikan, G. E., Hare, B., Snyder-Mackler, N., & MacLean, E. L. (2020). Estimating the heritability of cognitive traits across dog breeds reveals highly heritable inhibitory control and communication factors. *Animal Cognition, 23*(5), 953–964. https://doi.org/10.1007/s10071-020-01400-4

Gomes, A. C. R., Guerra, S., Silva, P. A., Marques, C., Trigo, S., Boogert, N. J., & Cardoso, G. C. (2020). Proactive common waxbills make fewer mistakes in a cognitive assay, the detour-reaching task. *Behavioral Ecology and Sociobiology, 74*(3). https://doi.org/10.1007/s00265-020-2809-2

Grueber, C. E., Nakagawa, S., Laws, R. J., & Jamieson, I. G. (2011). Multimodel inference in ecology and evolution: Challenges and solutions. *Journal of Evolutionary Biology, 24*(4), 699–711. https://doi.org/10.1111/j.1420-9101.2010.02210.x

Guillette, L. M., Hahn, A. H., Hoeschele, M., Przyslupski, A. M., & Sturdy, C. B. (2015). Individual differences in learning speed, performance accuracy and exploratory behaviour in black-capped chickadees. *Animal Cognition, 18*(1), 165–178. https://doi.org/10.1007/s10071-014-0787-3

Guillette, L. M., Naguib, M., & Griffin, A. S. (2017). Individual differences in cognition and personality. *Behavioural Processes, 134*, 1–3. https://doi.org/10.1016/j.beproc.2016.12.001

van Horik, J. O., Beardsworth, C. E., Laker, P. R., Langley, E. J. G., Whiteside, M. A., & Madden, J. R. (2019). Unpredictable environments enhance inhibitory control in pheasants. *Animal Cognition, 22*(6), 1105–1114. https://doi.org/10.1007/s10071-019-01302-0

van Horik, J. O., Beardsworth, C. E., Laker, P. R., Whiteside, M. A., & Madden, J. R. (2020). Response learning confounds assays of inhibitory control on detour tasks. *Animal Cognition, 23*(1), 215–225. https://doi.org/10.1007/s10071-019-01330-w

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). https://doi.org/10.1098/rspb.2018.0150

Isaksson, E., Utku Urhan, A., & Brodin, A. (2018). High level of self-control ability in a small passerine bird. *Behavioral Ecology and Sociobiology, 72*(7), 118. https://doi.org/10.1007/s00265-018-2529-z

Jelbert, S. A., Taylor, A. H., & Gray, R. D. (2016). Does absolute brain size really predict self-control? Hand-tracking training improves performance on the A-not-B task. *Biology Letters, 12*(2), 20150871. https://doi.org/10.1098/rsbl.2015.0871

Johnson-Ulrich, L., Holekamp, K. E., & Hambrick, D. Z. (2020). Innovative problem-solving in wild hyenas is reliable across time and contexts. *Scientific Reports, 10*(1), 13000. https://doi.org/10.1038/s41598-020-69953-5

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

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13). https://doi.org/10.18637/jss.v082.i13

Lacreuse, A., Gullstrand, J., & Fagot, J. (2016). Sex differences in inhibitory control in socially-housed baboons (*Papio papio*). *Behavioural Brain Research, 312*, 231–237. https://doi.org/10.1016/j.bbr.2016.06.032

Langley, E. J. G., Adams, G., Beardsworth, C. E., Dawson, D. A., Laker, P. R., Van Horik, J. O., Whiteside, M. A., Wilson, A. J., & Madden, J. R. (2020). Heritability and correlations among learning and inhibitory control traits. *Behavioral Ecology, 31*(3), 798–806. https://doi.org/10.1093/beheco/araa029

Lucon-Xiccato, T., & Bertolucci, C. (2020). Inhibitory control in zebrafish, *Danio rerio. Journal of Fish Biology, 97*(2), 416–423. https://doi.org/10.1111/jfb.14380

Lucon-Xiccato, T., Montalbano, G., & Bertolucci, C. (2020). Personality traits covary with individual differences in inhibitory abilities in 2 species of fish. *Current Zoology, 66*(2), 187–195. https://doi.org/10.1093/cz/zoz039

Macdonald, J. A., Beauchamp, M. H., Crigan, J. A., & Anderson, P. J. (2014). Age-related differences in inhibitory control in the early school years. *Child Neuropsychology, 20*(5), 509–526. https://doi.org/10.1080/09297049.2013.822060

MacLean, E. L., Hare, B., Nun, C. L., Addess, E., Amic, F., Anderson, R. C., Aureli, F., Baker, J. M., Bania, A. E., Barnard, A. M., Boogert, N. J., Brannon, E. M., Bray, E. E., Bray, J., Brent, L. J. N., Burkart, J. M., Call, J., Cantlo, J. F., Chek, L. G., … Zhao, Y. (2014). The evolution of self-control. *Proceedings of the National Academy of Sciences of the United States of America, 111*(20), E2140–E2148. https://doi.org/10.1073/pnas.1323533111

Martin, J. G. A., Nussey, D. H., Wilson, A. J., & Réale, D. (2011). Measuring individual differences in reaction norms in field and experimental studies: A power analysis of random regression models. *Methods in Ecology and Evolution, 2*(4), 362–374. https://doi.org/10.1111/j.2041–210X.2010.00084.x

McCune, K. B., Jablonski, P., Lee, S. I., & Ha, R. R. (2019). Captive jays exhibit reduced problem-solving performance compared to wild conspecifics. *Royal Society Open Science, 6*(1), 181311. https://doi.org/10.1098/rsos.181311

Mischel, W., Shoda, Y., & Rodriguez, M. L. (1989). Delay of gratification in children. *Science, 244*(4907), 933–938. https://doi.org/10.1126/science.2658056

Moeller, F. G., Barratt, E. S., Dougherty, D. M., Schmitz, J. M., & Swann, A. C. (2001). Psychiatric aspects of impulsivity. *American Journal of Psychiatry, 158*(11), 1783–1793. https://doi.org/10.1176/appi.ajp.158.11.1783

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

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., Hamblin, S., Cole, E. F., Aplin, L. M., & Quinn, J. L. (2015). Taking the operant paradigm into the field: Associative learning in wild great tits. *PLoS One, 10*(8), Article e0133821. https://doi.org/10.1371/journal.pone.0133821

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 of the United States of America, 108*(38), 15898–15903. https://doi.org/10.1073/pnas.1111560108

Morand-Ferron, J., & Quinn, J. L. (2015). The evolution of cognition in natural populations. *Trends in Cognitive Sciences, 19*(5), 235–237. https://doi.org/10.1016/j.tics.2015.03.005

Mouchet, A., & Dingemanse, N. J. (2021). A quantitative genetics approach to validate lab- versus field-based behavior in novel environments. *Behavioral Ecology, 32*(5), 903–911. https://doi.org/10.1093/beheco/arab059

Muth, F., Cooper, T. R., Bonilla, R. F., & Leonard, A. S. (2018). A novel protocol for studying bee cognition in the wild. *Methods in Ecology and Evolution, 9*(1), 78–87. https://doi.org/10.1111/2041–210X.12852

Naimi, B., Hamm, N. A. S., Groen, T. A., Skidmore, A. K., & Toxopeus, A. G. (2014). Where is positional uncertainty a problem for species distribution modelling? *Ecography, 37*(2), 191–203. https://doi.org/10.1111/j.1600-0587.2013.00205.x

Nakagawa, S., Johnson, P. C. D., & Schielzeth, H. (2017). The coefficient of determination R2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of The Royal Society Interface, 14*(134), 20170213. https://doi.org/10.1098/rsif.2017.0213

Nakagawa, S., & Schielzeth, H. (2010). Repeatability for Gaussian and non-Gaussian data: A practical guide for biologists. *Biological Reviews, 85*(4), 935–956. https://doi.org/10.1111/j.1469–185X.2010.00141.x

O'Shea, W., O'Halloran, J., & Quinn, J. L. (2018). Breeding phenology, provisioning behaviour, and unusual patterns of life history variation across an anthropogenic heterogeneous landscape. *Oecologia, 188*(4), 953–964. https://doi.org/10.1007/s00442-018-4155-x

O'Shea, W., Serrano-Davies, E., & Quinn, J. L. (2017). Do personality and innovativeness influence competitive ability? An experimental test in the great tit. *Behavioral Ecology, 28*(6), 1435–1444. https://doi.org/10.1093/beheco/arx104

Quinn, J. L., Patrick, S. C., Bouwhuis, S., Wilkin, T. A., & Sheldon, B. C. (2009). Heterogeneous selection on a heritable temperament trait in a variable environment. *Journal of Animal Ecology, 78*(6), 1203–1215. https://doi.org/10.1111/j.1365-2656.2009.01585.x

R Core Team. (2017). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. http://www.R-project.org/.

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

Reddy, R. B., MacLean, E. L., Sandel, A. A., & Hare, B. (2015). Social inhibitory control in five lemur species. *Primates, 56*(3), 241–252. https://doi.org/10.1007/s10329-015-0467-1

Reichert, M. S., Crofts, S. J., Davidson, G. L., Firth, J. A., Kulahci, I. G., & Quinn, J. L. (2020). Multiple factors affect discrimination learning performance, but not between-individual variation, in wild mixed-species flocks of birds. *Royal Society Open Science, 7*(4), 192107. https://doi.org/10.1098/rsos.192107

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

Santacà, M., Busatta, M., Savaşçı, B. B., Lucon-Xiccato, T., & Bisazza, A. (2019). The effect of experience and olfactory cue in an inhibitory control task in guppies, *Poecilia reticulata. Animal Behaviour, 151*, 1–7. https://doi.org/10.1016/j.anbehav.2019.03.003

Savaşçı, B. B., Lucon-Xiccato, T., & Bisazza, A. (2021). Ontogeny and personality affect inhibitory control in guppies, *Poecilia reticulata. Animal Behaviour, 180*, 111–121. https://doi.org/10.1016/j.anbehav.2021.08.013

Shaw, R. C. (2017). Testing cognition in the wild: Factors affecting performance and individual consistency in two measures of avian cognition. *Behavioural Processes, 134*, 31–36. https://doi.org/10.1016/j.beproc.2016.06.004

Shaw, R. C., Boogert, N. J., Clayton, N. S., & Burns, K. C. (2015). Wild psychometrics: Evidence for 'general' cognitive performance in wild New Zealand robins, *Petroica longipes. Animal Behaviour, 109*, 101–111. https://doi.org/10.1016/j.anbehav.2015.08.001

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

Soha, J. A., Peters, S., Anderson, R. C., Searcy, W. A., & Nowicki, S. (2019). Performance on tests of cognitive ability is not repeatable across years in a songbird. *Animal Behaviour, 158*, 281–288. https://doi.org/10.1016/j.anbehav.2019.09.020

Sonnenberg, B. R., Branch, C. L., Pitera, A. M., Bridge, E., & Pravosudov, V. V. (2019). Natural selection and spatial cognition in wild food-caching mountain chickadees. *Current Biology, 29*(4), 670–676.e3. https://doi.org/10.1016/j.cub.2019.01.006

Thornton, A., Isden, J., & Madden, J. R. (2014). Toward wild psychometrics: Linking individual cognitive differences to fitness. *Behavioral Ecology, 25*(6), 1299–1301. https://doi.org/10.1093/beheco/aru095

Toledo, S., Shohami, D., Schiffner, I., Lourie, E., Orchan, Y., Bartan, Y., & Nathan, R. (2020). Cognitive map-based navigation in wild bats revealed by a new high-throughput tracking system. *Science, 369*(6500), 188–193. https://doi.org/10.1126/science.aax6904

Troisi, C. A., Cooke, A. C., Davidson, G. L., de la Hera, I., Reichert, M. S., & Quinn, J. L. (2021). No evidence for cross-contextual consistency in spatial cognition or behavioral flexibility in a passerine. *Animal Behavior and Cognition, 8*(3), 446–461. https://doi.org/10.26451/abc.08.03.08.2021

Vernouillet, A. A. A., Stiles, L. R., Andrew McCausland, J., & Kelly, D. M. (2018). Individual performance across motoric self-regulation tasks are not correlated for pet dogs. *Learning & Behavior, 46*(4), 522–536. https://doi.org/10.3758/s13420-018-0354-x

Völter, C. J., Tinklenberg, B., Call, J., & Seed, A. M. (2018). Comparative psychometrics: Establishing what differs is central to understanding what evolves. *Philosophical Transactions of the Royal Society B: Biological Sciences, 373*(1756), 20170283. https://doi.org/10.1098/rstb.2017.0283

Wilkinson, L. (2011). ggplot2: Elegant graphics for data analysis. *Biometrics, 67*(2), 678–679. https://doi.org/10.1111/j.1541-0420.2011.01616.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

Appendix: Detour Task in the Wild

Methods

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

During the training phase, the opaque barrier was inverted so that it covered the nest hole (Fig. 1a). This phase ensured birds could perform and were familiar with the motor action of going under the barrier. Birds were visually observed from the camcorder footage, and if they went under the barrier without touching it a minimum of three times over the course of the training phase, they were advanced to the test phase the following day. During the test phase, the barrier was transparent (i.e. without camouflage tape on the plastic film; Fig. 1b). The test phase was completed once the birds had made a minimum of five attempts to enter the nest hole. These attempts, hereafter referred to as 'trials', were defined as either colliding with or touching the transparent barrier or accessing the nest hole under the barrier without touching it. A minimum of five trials was set as the criterion as this was the lowest previsioning rate (i.e. five visits within 1 h) recorded in our population (W. O'Shea, personal communication, 2016). However, some birds did not achieve five trials in a 1 h experiment, and therefore the test phase was repeated the following day. For welfare reasons associated with nest disturbance, we allowed up to 6 days of experiments across all phases.

Results

One of the birds for which we had repeat measures was recorded in year 1, year 2 and in captivity. To control for potential effects of memory, we reran the contextual repeatability model excluding this bird's performance in year 2. Nonadjusted contextual repeatability in performance across tasks was low but significant ($R = 0.17$, $P < 0.001$, CI = 0.13, 0.18), and remained significant when controlling for sex (adjusted $R = 0.16$, $P < 0.001$, CI = 0.12, 0.17).

Table A1
List of the 10 study sites in the Bandon Valley area, co. Cork (Ireland), where birds were studied

Study site	Habitat type	Coordinates (DD)
Ballinphelic	Coniferous	51.840074, -8.628249
Castlebernard*	Mixed	51.741821, -8.773443
Piercetown	Coniferous	51.789310, -8.451774
Dukes Wood*	Mixed	51.787396, -8.755946
Farran	Coniferous	51.700448, -8.805728
Garretstown	Coniferous	51.654770, -8.616293
Inishannon*	Mixed	51.763048, -8.662415
Kilbrittain*	Mixed	51.671880, -8.683837
Lissarda	Coniferous	51.858086, -8.859128
Shipool‡	Mixed	51.737620, -8.631437

We show the coordinates in decimal degrees (DD) and the main habitat type (mixed deciduous forest versus coniferous plantation) for each site.
*Sites that were included in year 2.
‡ A site where birds were not caught for the captive task.

Table A2
Test statistics from the global binomial GLMMs of detour performance

	Estimate ± SE	Z	P
Wild detour task			
Intercept	1.1±0.43	2.56	0.01
Brood size	-0.02±0.19	-0.09	0.93
Total training trials	-0.05±0.17	-0.3	0.76
Year (Year 1)	0.23±0.43	0.53	0.59
Wing length	0.36±0.27	1.33	0.18
Lay date	-0.35±0.2	-1.77	0.08
Sex (Female)	-1.19±0.52	-2.31	0.02
Captive detour task			
Intercept	-0.34±0.27	-1.3	0.19
Sex (Female)	-0.18±0.26	-0.68	0.5
Age (Juvenile)	0.16±0.28	0.57	0.57
Time to complete test	0.11±0.12	0.94	0.35
Total rewards eaten	0.18±0.13	1.36	0.17
Exploration behaviour	0.17±0.14	1.25	0.21
Habitat (Conifer)	0.11±0.36	0.312	0.76

Performance was measured as the proportion of successful test trials for the wild detour task (N=84) and the captive detour task (N=35). All continuous variables were scaled. Reference categories for binary variables are in parentheses.
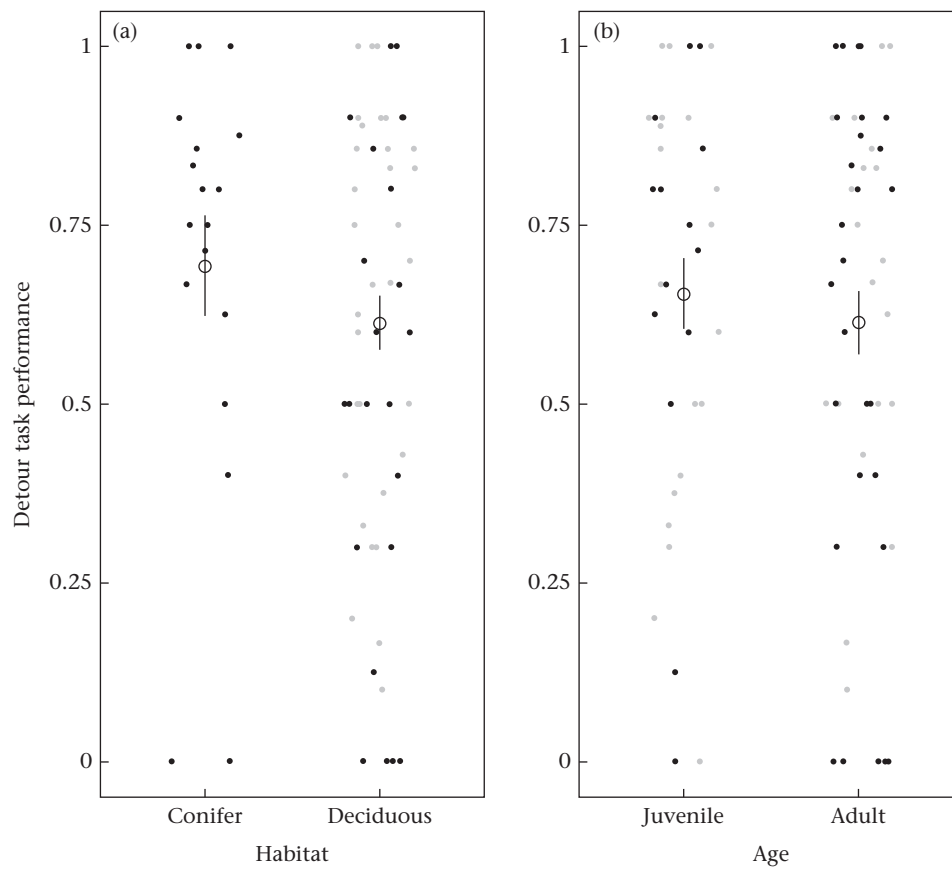
Figure A1. Detour task performance (the proportion of successful trials out of total trials, where higher values indicate better performance) across (a) habitats and (b) age, measured in the wild. Points represent individuals: black points were taken in year 1 and grey in year 2. Note that no individuals were measured in conifer habitats in year 2. Data points are jittered along the horizontal axis to reduce overlap. Open circle and line represent mean ± SE across both years.
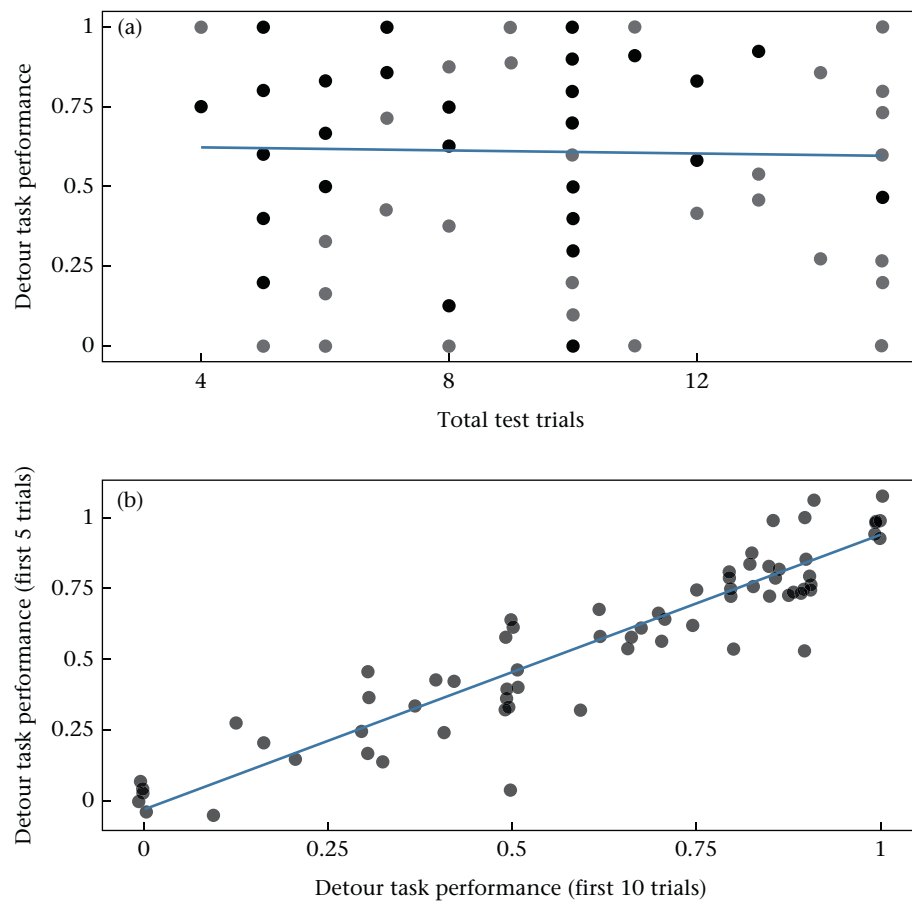
Figure A2. The number of test trials and detour task performance in the wild (the proportion of successful trials out of total trials, where higher values indicated better performance). (a) Individuals' overall performance in relation to the number of test trials. (b) Individuals' overall performance calculated from the first 10 trials in relation to their overall performance calculated from the first five trials.
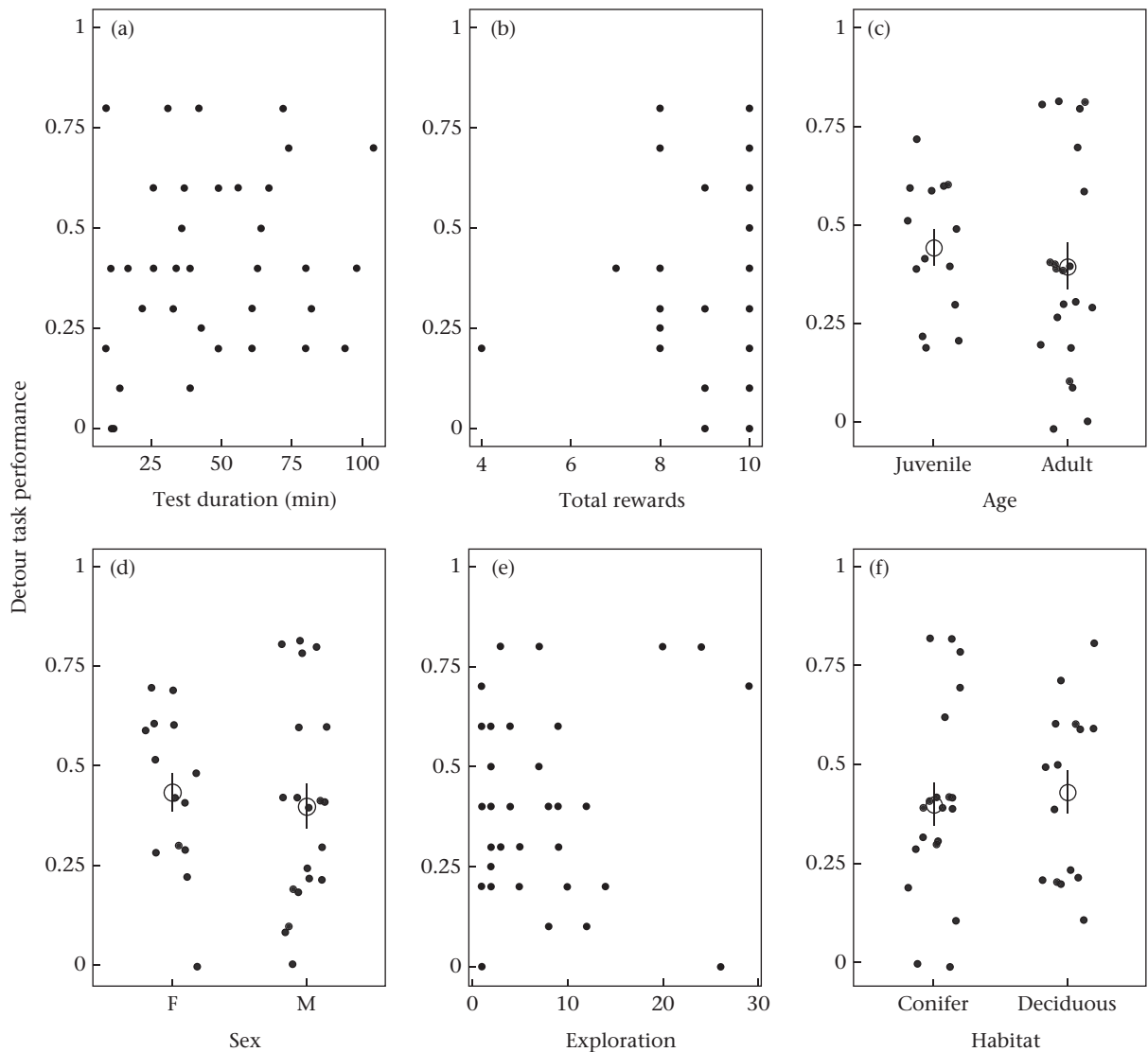
Figure A3. Detour task performance (the proportion of successful trials out of total trials, where higher values indicate better performance) in captivity in relation to (a) test duration, (b) total rewards eaten, (c) age, (d) sex (F=female, M=male), (e) exploration behaviour (the sum of the number of hops and flights) and (f) habitat. Points represent individuals and are jittered along the horizontal axis to reduce overlap in (c), (d) and (f). Open circle and line represent mean ± SE. See Table 1 for model results.