**Author for correspondence:**
Ella McCallum
e-mail: ella-mccallum@outlook.com

THE ROYAL SOCIETY PUBLISHING

# Repeatability and heritability of inhibitory control performance in wild toutouwai (*Petroica longipes*)

## Ella McCallum and Rachael C. Shaw

School of Biological Sciences, Te Herenga Waka Victoria University of Wellington, Wellington, New Zealand

EM, 0009-0002-6236-4611; RCS, 0000-0002-4831-8283

Despite increasing interest in the evolution of inhibitory control, few studies have examined the validity of widespread testing paradigms, the long-term repeatability and the heritability of this cognitive ability in the wild. We investigated these aspects in the inhibitory control performance of wild toutouwai (North Island robin; *Petroica longipes*), using detour and reversal learning tasks. We assessed convergent validity by testing whether individual performance correlated across detour and reversal learning tasks. We then further evaluated task validity by examining whether individual performance was confounded by non-cognitive factors. We tested a subset of subjects twice in each task to estimate the repeatability of performance across a 1-year period. Finally, we used a population pedigree to estimate the heritability of task performance. Individual performance was unrelated across detour and reversal learning tasks, indicating that these measured different cognitive abilities. Task performance was not influenced by body condition, boldness or prior experience, and showed moderate between-year repeatability. Yet despite this individual consistency, we found no evidence that task performance was heritable. Our findings suggest that detour and reversal learning tasks measure consistent individual differences in distinct forms of inhibitory control in toutouwai, but this variation may be environmentally determined rather than genetic.

## 1. Introduction

Individuals often exhibit striking differences in their ability to suppress prepotent responses, referred to as inhibitory control [1–6]. Deficits in this executive function are associated with impulsive acts and reliance on outdated information or irrelevant stimuli, which can prevent an individual from reaching their

goals [7,8]. In humans, superior inhibition is linked to better health, social and financial outcomes [9]. Likewise, this ability is thought to facilitate many fitness-relevant behaviours in non-human animals. For example, researchers have speculated that inhibition may allow animals to resist launching a premature attack on prey [10], cache food for later [11,12] or share resources with conspecifics [13]. The last decade has seen an upsurge of interest in the causes and consequences of individual variation in this ability [1,13–16]. However, to understand whether and how selection acts on inhibitory control, it is crucial to assess whether tasks used to assess inhibitory control are valid measures of repeatable, heritable individual differences in this ability in the wild [17,18].

In non-human animals, two of the most frequently assessed forms of inhibitory control are motor and cognitive inhibition. Motor inhibition describes the ability to withhold prepotent motor actions [8,19]. It is often measured using detour tasks, where individuals must inhibit reaching directly for a reward through a transparent barrier and instead detour to an opening [20]. By contrast, cognitive inhibition involves the suppression of unwanted thoughts and memories, including 'intentionally forgetting' information that is no longer accurate [7,21]. This can be assessed with reversal learning tasks, where animals learn to respond differently to different stimuli, then must inhibit acting on this association and learn a new one when reward contingencies are reversed [22]. Reversal learning tasks also measure cognitive flexibility, an executive function that allows animals to adjust to changing task demands and priorities [22,23]. Inhibitory control is a core component of cognitive flexibility, as modifying thoughts and behaviour requires the inhibition of a previous perspective or response [7]. As detour and reversal learning tasks require little equipment and can be adapted to a range of species, they are widely used in captivity and the field [1,3–5]. However, there are concerns that these tasks may not provide accurate assays of inhibitory control [3,4].

In a well-designed cognitive test battery, an individual's performances should be highly correlated across tasks measuring the same cognitive construct, a concept known as convergent validity [24]. Assessing convergent validity can indicate whether tasks measure the expected cognitive abilities and provide deeper insight into the structure of cognition [25]. Studies of non-human animals have often neglected convergent validity by using a single task or focusing on group-level performance rather than individual differences [25]. Where it has been assessed, individual performances have often shown no correlation across putative inhibitory control tasks [2,25–29], suggesting that these measure distinct abilities which are not underpinned by a 'general' inhibitory control ability. Understanding the structure of inhibitory control is vital to understand its evolution. If inhibitory control is best understood as a collection of independent abilities with different roles in guiding behaviour, then examining the fitness correlates of performance in a single task may result in misleading conclusions.

A lack of convergent validity can also arise if tasks measure unwanted variables [25]. Behavioural tasks can be confounded by non-cognitive factors which vary across species and task type. For example, pheasants (*Phasianus colchicus*) performed better on a detour task when they had experience with transparent objects, but body condition did not affect performance [3]. By contrast, toutouwai (*Petroica longipes*) in better condition—but not those with experience—performed better on detour task [4]. Personality, defined as individual behavioural differences that are consistent across time and context [30], can also influence cognitive task performance in unpredictable ways [31]. Higher levels of boldness have been linked to better detour performance [32,33] but worse reversal learning ability [34]. Bolder, neophilic and/or exploratory individuals may also be more likely to participate in voluntary tests or be less fearful and consequently perform better [35,36]. Investigating which factors contribute to task performance can improve the validity of measures [17,37]. Possible confounds can then be accounted for within tasks (e.g. by habituating animals to the apparatus) or by measuring variables in separate assays and controlling for them statistically [17].

Another crucial, but often neglected, step in validating cognitive tasks is assessing whether they measure consistent individual differences in performance [25,38]. Repeatability describes the proportion of phenotypic variation due to between-subject variation [39]. Cognitive repeatability has typically been measured over short timescales in captivity, which can bias results due to similarity in the genetic and environmental influences on performance [38,40]. Only a small number of studies have taken repeated measures of inhibitory control performance over long timescales in the wild. Detour task performance showed significant (albeit low) long-term repeatability in great tits (*Parus major*) [6], but not in toutouwai (*Petroica longipes*) [4]. Australian magpies (*Gymnorhina tibicen*) showed moderate long-term repeatability in a reversal learning task, but no significant repeatability in a detour task [41] (despite high short-term repeatability in both tasks [5]). Understanding the long-term consistency of inhibitory control in the wild is crucial, as the ability to interpret links between cognition and other traits such as personality, behaviour and fitness depends on whether an

individual's cognition is temporally stable [42]. Moreover, repeatability can indicate the upper limit of heritability [39,43], a measure that describes the proportion of trait variation that is driven by genetic differences [44].

Exploring the genetic basis of cognitive variation in natural populations is crucial, as selection can only operate on heritable traits, and the speed of evolution partly depends on the strength of this heritability [43]. However, heritability estimates can be logistically difficult to obtain from natural populations [43]. While repeatability is a valuable substitute for heritability, it does not separate genetic and environmental contributions to individual variation [39,43]. A more reliable way to assess a trait's evolutionary potential is to measure narrow-sense heritability by testing related individuals and calculating the proportion of additive genetic variation [39]. The heritability of inhibitory control is well-studied in humans [45], showing a moderate to strong genetic component [45–48], but has received less attention in non-human animals, with estimates ranging from low and non-significant [49,50] to moderate [51–53]. To date, the heritability of cognitive traits has largely been assessed using laboratory populations of model species [18,54] which may not represent the genetic variation of their wild counterparts [55–58].

In this study, we assessed the validity, repeatability and heritability of inhibitory control task performance in wild toutouwai (North Island robin; *Petroica longipes*), a small, insectivorous passerine endemic to New Zealand. Both sexes cache food and males share food with their mates during the breeding season [59]. As both behaviours likely require individuals to overcome a prepotent impulse to eat food immediately, inhibitory control should be ecologically relevant to this species. We used detour reaching and reversal learning tasks to quantify individual differences in inhibitory control. We examined convergent validity by exploring how individual performances correlated across task type. To further assess task validity, we investigated whether boldness, body condition or prior experience confounded performance. We tested a subset of birds twice in each task to determine whether performance was repeatable across years. Finally, we measured the heritability of task performance using a population pedigree.

# 2. Methods

We conducted this study at Zealandia Ecosanctuary, a 225 ha fenced sanctuary in Wellington, New Zealand (41°18′ S, 174°44′ E). Subjects were colour-banded birds belonging to our long-term research population [60] and ranged from six months to 12 years of age. Toutouwai in Zealandia Ecosanctuary are protected from non-native mammalian predators by a predator exclusion fence but are exposed to native avian predators and natural weather conditions. Members of our research population are given small numbers of mealworms (*Tenebrio molitor* larvae) during experiments and breeding season monitoring, but do not otherwise receive supplementary feeding or any other management interventions. Testing took place from April to September in 2021 and 2022. Each year, we tested toutouwai in a detour task, an associative learning task, and a reversal learning task. Sample sizes are stated below for each task, with a detailed breakdown provided in electronic supplementary material, table S1.

Experiments were conducted between 9.00 and 15.30, with a consistent testing location set up in a flat, central area of a subject's non-breeding territory. Subjects received up to 20 trials per day, with a freshly killed mealworm used as the reward in all trials. We conducted test sessions on consecutive calendar days, but if this was infeasible (e.g. due to severe weather), testing continued on the next possible day. After completing a trial, birds could begin the next trial as soon as the experimenter had re-baited the apparatus. If a bird did not participate within 2 min in three consecutive trials, testing stopped and was resumed on the next testing day, as this suggested a lack of motivation. An exception to this was if a bird was engaged in important activities such as territory defence. In these cases, the session continued once it returned from the activity.

## 2.1. Detour task

Between April and May 2021, we tested motor inhibition in 45 birds with a cylinder detour task used previously for this species [4]. The apparatus was a plastic cylinder (4 cm diameter × 5 cm length) wrapped in black tape during habituation and training (figure 1*a*), and transparent during testing (figure 1*b*), with a mealworm placed in the centre. A white plastic barrier ensured that subjects first encountered the cylinder face, rather than directly approaching an open end.

**Figure 1.** Apparatuses used for cognitive tasks: (*a*) opaque cylinder used for habituation and training in the cylinder detour task, (*b*) transparent cylinder used for the test phase in the cylinder detour task, (*c*) opaque barrier used for habituation and training in the U-barrier detour task, (*d*) transparent barrier used for the test phase in the U-barrier detour, and (*e*) two tiles with removable lids used for the spatial association and reversal learning tasks.

We began with a habituation procedure to minimize the effect of neophobia [61]. Here, the opaque cylinder was placed immediately in front of the white plastic barrier so subjects could readily see the mealworm. To pass habituation, a bird needed to take the reward within 2 min in three consecutive trials. Immediately after passing habituation, we began the training phase. During training, the opaque cylinder was placed inside the barrier (figure 1*a*) and the bird learned to reach its head between the barrier wall and cylinder end to retrieve the reward. This ensured that failures during the subsequent test phase were not merely caused by the bird not knowing the detour movement [20]. To pass training, a bird needed to retrieve the mealworm within 2 min in four consecutive trials without first pecking at the cylinder wall.

The test phase began immediately after a bird passed training. The opaque cylinder was replaced with a transparent cylinder (figure 1*b*) so that the mealworm was now visible and subjects could peck

directly towards it [20]. To pass the test, a bird needed to inhibit this response and retrieve the mealworm without first pecking at the cylinder in 6/7 consecutive trials (either within or across testing sessions). If a bird did not take the reward within 2 min and did not peck the cylinder, this trial did not contribute to their score, as failure to interact with the apparatus entirely may show that a bird is unmotivated or fearful but does not indicate poor inhibitory control. All individuals who participated in the test phase reached the 6/7 criterion, and their performance score was the number of trials required to achieve this.

From April to September 2022, we tested 45 birds in a second detour task. Thirty-one of these individuals had participated in the cylinder detour task. The apparatus was a plastic U-shaped barrier (9 cm height × 6 cm diameter) mounted on a circular metal base. It was wrapped in black tape during habituation and training (figure 1c) and transparent during the test phase (figure 1d). A bird needed to detour around the U-barrier to retrieve a mealworm placed in the centre. To prevent the bird from directly approaching the open side of the U-barrier, the experimenter sat facing the open side and only started a trial when the bird was positioned so that they would first encounter the barrier face when moving directly towards the reward. This task invokes the same putative cognitive ability as the cylinder detour, but involves a locomotor detour (moving the whole body around the barrier) rather than the reaching detour (reaching head or limb around the barrier) [20]. We changed the visual appearance and required motor action to reduce the likelihood that prior experience influenced U-barrier task performance, while also allowing us to assess how barrier shape affects detour performance, which has rarely been tested.

During habituation, we placed a single mealworm immediately beside the opaque U-shaped barrier (left/right side order was randomly selected) for the first two trials and behind the barrier in trial three. To pass habituation, a bird needed to retrieve the worm from these locations, each within 2 min. If the bird did not pass a trial, we removed the worm then placed it back in the same location to regain their interest, then began a new trial. Training began immediately after a bird passed habituation, with the mealworm placed inside the opaque U-barrier (figure 1c). To pass training, the bird needed to detour around the barrier and retrieve the worm within 2 min in four consecutive trials, without pecking the barrier wall. All birds passed habituation and training.

The test phase began immediately after the bird passed training. The set-up was identical to the training phase except that the barrier was now transparent (figure 1d). A bird needed to retrieve a mealworm from behind the barrier without first pecking the barrier wall in 6/7 consecutive trials either within or across test sessions. As with the cylinder task, time-outs without pecking were not counted as failed trials and the performance score was the number of trials a bird required to reach the 6/7 criterion.

## 2.2. Spatial association and reversal learning task

In both 2021 and 2022, we assessed cognitive inhibition using a spatial associative learning task ($N_{2021} = 41$; $N_{2022} = 37$), followed by a spatial reversal learning task ($N_{2021} = 37$; $N_{2022} = 35$). Testing took place between July and September in 2021, and June and August in 2022. Twenty-five birds were tested in both years of the association task, and 21 birds in both reversal learning tasks. In both years, testing began after subjects completed the detour task. For both tasks, the apparatus consisted of two 4 × 4 cm plastic tiles with a 2 cm diameter by 0.8 cm deep well drilled in the centre. These were placed 12 cm apart on a wooden board and separated by three white plastic barriers (21 cm length × 10 cm height). Tiles were covered by brown laminated paper lids (5 cm diameter) to conceal the wells (figure 1e).

Prior to testing, we used a behavioural shaping procedure to train subjects to remove lids and retrieve hidden mealworms. In level one, a lid leaned against the tile and the mealworm was visible. In level two, a lid was folded in half so that it formed a tent shape over the tile well and the mealworm was only partially visible in the gap between the tile and lid. In level three, a flat lid covered the tile well and hid the mealworm. To pass each level, a toutouwai needed to take the mealworm in three consecutive trials within 2 min. This necessitated removing the lid in levels two and three. If a bird failed a trial, they regressed to the previous level and needed to pass this again.

For the association learning task, toutouwai learned to associate a spatial location (either the left or right side of the testing apparatus) with a reward. We first gave a bird a spatial preference test and probe trial. In the preference test, the wells of both tiles were baited, and a bird could remove both lids. The location that a bird selected second (i.e. their less preferred location) was then rewarded during their association trials. Next, we gave a bird a probe trial where only one location was baited, and they

could remove both lids. As the spatial preference test had rewarded both locations, the probe trial allowed birds to learn that only one location was now rewarded, without this contributing to their overall score. Following the probe trial, they began the association task. Here, only one location was rewarded, and a bird could only remove one lid per trial. Subjects had to choose the rewarded lid on 10/12 consecutive test trials within or across sessions to pass the task (exceeding the chance expectation of 6/12 trials correct, two-tailed binomial test: $p = 0.039$). Birds were given 100 trials to reach this criterion on the association task and 120 trials on the reversal task. The task performance score was the number of trials a bird took to reach the 10/12 criterion. In 2021, two females did not reach criterion after completing 100 trials and so did not move on to the reversal learning task. However, they were given performance scores of 100 and included in analyses of this task, as despite not being an exact score, this indicated that they performed substantially worse than average (see results for the mean number of trials to reach criterion). We also ran all analyses excluding these birds that did not reach criterion, but this did not change the overall results, so we retained them for maximum statistical power. All birds reached criterion in 2022.

The reversal learning task began the day after a bird completed the association learning task. Typically, this was the next calendar day, but this was not always possible. Potentially, individuals with a longer interval between these tasks may have a weaker memory of the association reward location [62], and thus have less difficulty reversing this association. Therefore, to quantify how well birds recalled the rewarded location they had learned in the association task, all birds received three association refresher trials prior to starting the reversal learning task. Here, the mealworm was placed in the same well rewarded in the association learning task. Immediately following this refresher, we began the reversal learning task. For this, the previously rewarded spatial location was now unrewarded and vice versa. Again, a bird was given a probe trial to learn that only one position was rewarded. Subjects then needed to choose the correct location in 10/12 consecutive test trials to pass. The performance score was the number of trials to reach criterion. Four individuals (two per year) did not reach criterion after completing 120 trials. They were given performance scores of 120 and included in analyses. As with the associative learning task, their exclusion did not change the overall results.

## 2.2.1. Factors affecting task performance

Despite familiarizing subjects with the cognitive tasks during habituation and training, some individuals potentially remained fearful of interacting with the apparatuses and experimenter. Therefore, we tested whether boldness—defined as behavioural responses to risky, but not novel, situations [63]—affected performance during the test phase.

To quantify boldness, we measured subjects' latencies to retrieve a mealworm close to a human experimenter. Neither the food nor experimenter was novel, but the context could still be perceived as risky. To begin a boldness test, the experimenter threw a single mealworm to approximately 1 m in front of their own feet to standardize start distance as much as possible. Immediately after birds ate this mealworm, the experimenter placed a second, freshly killed mealworm 30 cm from their own feet onto a cleared patch of ground. Latency to retrieve this worm was taken as the boldness measure, with a faster latency indicating a bolder individual. In each year, subjects received two boldness tests on consecutive testing days (typically the next calendar day). Twenty-five subjects were tested across both years (2021 and 2022), allowing us to assess whether latency to retrieve the mealworm was temporally repeatable. Significant repeatability would provide evidence that this experiment captured boldness, as behaviour must be consistent across time and context to be considered personality [63]. In 2021, 33 of our subjects received boldness tests in April prior to the detour task, and 11 were tested in July following the detour task. In 2022, subjects ($N = 37$) received boldness tests during July and August following the reversal task. Latencies were coded *in situ* in 2021 and using BORIS software [64] in 2022.

To examine the possibility that latency to retrieve the mealworm during boldness tests may largely measure hunger rather than boldness, birds were given a mealworm on a scale following boldness test sessions in 2022. We divided their mass (measured to the nearest 0.1 g) by their tarsus measurement (mm) to obtain their body condition [4,65]. We also calculated body condition during all cognitive tasks. Again, toutouwai were given a mealworm on a scale after test sessions, and we obtained a single value for body condition in each task by dividing a subject's average mass across all sessions by their tarsus measurement [4,65]. We obtained age data from breeding records. Where subjects were banded as adults (i.e. immigrants or those born before monitoring began in 2014), banding year was used as a proxy for birth year. We determined sex using breeding records, plumage coloration (mature toutouwai are sexually dimorphic [66]), and vocalizations (only males give full-song [67]).

## 2.3. Statistical analysis

We performed statistical analyses in R software [68], using the packages 'glmmTMB' [69] to run models, 'MuMIn' [70] for model averaging, 'rptR' [71] for repeatability estimates, 'MCMCglmm' [72] for heritability models, 'Hmisc' [73] for correlation matrices and 'Factoextra' [74] for principal components analysis (PCA). We follow guidelines from Harper [75] for categorizing the strength of repeatability estimates on a scale from slight ($R < 0.2$) to very high ($R > 0.9$) repeatability.

### 2.3.1. Relationships between tasks

To explore whether detour and reversal learning tasks demonstrated convergent validity, we used Spearman rank correlation and PCA to compare performances across tasks for individuals who completed all three cognitive tasks during each year ($N_{2021} = 37$, $N_{2022} = 35$). We ran a separate Spearman rank correlation matrix and PCA for task performance in each year, so that only one replicate of each task was included for each bird per year. Only principal components with eigenvalues $> 1$ were retained in the PCA [76,77]. If individual performances in the detour and reversal task show a significant, positive correlation and load strongly in the same direction onto the same principal component (while the associative learning task loads comparatively weakly or in the opposite direction) this would provide evidence that tasks have convergent validity [2,25].

### 2.3.2. Factors affecting cognitive task performance

We used linear mixed model-based repeatability analysis to determine whether latency to retrieve a mealworm under threat was temporally repeatable. Latencies were log-transformed to ensure normality. The model included both years of data, resulting in a total of 162 boldness measures across 56 individuals. As individuals were measured twice within each year, we included testing year (2021 or 2022) as a fixed factor to calculate adjusted repeatability. The model also included bird ID as a random factor.

To investigate whether latency to retrieve the mealworm during boldness tests was confounded by hunger, we used a linear mixed model to analyse the effect of body condition on latency to retrieve the mealworm during the 2022 boldness tests ($N = 51$ observations across 32 individuals: not all individuals remained on the scale long enough for a mass measurement). Bird ID was included as a random factor.

We used model averaging to analyse the factors affecting task performance. We expected that non-cognitive factors could have different effects on the U-barrier detour and cylinder detour tasks due to their different apparatuses and required motor actions. For instance, boldness could be more influential for the cylinder task, as birds needed to insert their head in a narrow space between the cylinder and surrounding shield. As our sample size was too small to analyse both detour tasks together and fit interaction terms to account for this, we analysed the factors affecting performance in each detour task (i.e. each year) separately using negative binomial generalized linear models (GLMs) as Poisson models were overdispersed. The association and reversal learning models included both years of task performance as the tasks were identical. We analysed associative learning data and reversal learning data using negative binomial generalized linear mixed models (GLMMs) as Poisson models were overdispersed. We fitted bird ID as a random factor in both GLMMs.

In all models, we included task performance as the response variable, with boldness, body condition, age and sex included as fixed factors. The boldness measure was an individual's average latency to retrieve the mealworm across the two testing days during the year they participated in the task. In the reversal learning task model, we also included the number of refresher trials (out of three) that an individual passed as a proxy for memory of the association reward location. As tarsus measurements, sex or boldness scores were not available for all individuals, we analysed the factors affecting performance for 37 birds in the cylinder detour task, 36 birds in the U barrier detour task, 48 birds for the spatial associative learning task and 46 birds in the spatial reversal learning task.

For the models of the association and reversal learning tasks, we included year as a fixed factor. To test whether prior experience impacted performance, we also ran Mann–Whitney $U$-tests to assess whether experienced birds performed better than naïve birds during the 2022 association and reversal tasks. Similarly, in the U-barrier detour model we included whether an individual was tested in the cylinder task as a fixed factor to assess the effect of prior experience. For each global model described above, we constructed models containing all possible combinations of the predictors. We then

averaged parameters across the full model set and calculated 95% CIs and the RVI for each mean $\beta$. We used model averaging across the full set of possible models of task performance rather than making inferences from a single model, as this can account for model uncertainty and ensure that estimates for fixed effects are robust to different model specifications [78].

### 2.3.3. Repeatability and heritability of task performance

We assessed the repeatability of detour, association and reversal learning task performance between years by measuring the amount of variation in task performance attributable to differences between-individuals [79]. We ran three separate Poisson GLMM-based repeatability models with task performance as the response variable and bird ID as a random factor. We included repeated measures for birds tested in both years ($N_{detour} = 62$ observations across 31 individuals; $N_{association} = 50$ observations across 25 individuals; $N_{reversal} = 42$ observations across 21 individuals). To increase power, we also included single measures for birds tested in one year only ($N_{detour} = 28$, $N_{association} = 28$, $N_{reversal} = 30$) [80].

We estimated the narrow-sense heritability of task performances using Bayesian GLMM-based animal models. Animal models use a population pedigree to estimate the proportion of additive genetic variance behind a trait. As toutouwai are monogamous [81] and our population has been monitored since 2014 [60], pedigree information was available without genetic analysis. Our pedigree (see the electronic supplementary material) included all subjects tested in our cognitive tasks. We included mother and father ID for test subjects hatched within the study site after monitoring began in 2014. Immigrants and birds hatched before 2014 were included in the pedigree despite unknown parentage, as unlike parent–offspring regression and half-sibling heritability designs, animal models take into account all forms of relatedness between individuals, maximizing statistical power [82]. Animal models can also account for environmental and genetic biases on heritability estimates [82,83].

We ran four models: associative learning task (2021 and 2022), reversal learning task (2021 and 2022), cylinder task and detour task. All models included task performance as the response variable and the following random effects: bird ID linked to the pedigree, bird ID independent of pedigree (to account for permanent environmental effects, which are environmental effects on an individual's phenotype that are constant across repeated measures [83]), mother and father ID (to account for parental effects on offspring phenotype independent of additive genetic effects [84]) and a dominance matrix calculated using the 'nadiv' package [85] (to account for genetic dominance effects, which can contribute to non-additive genetic variance [86]). From these models, we estimated $h^2$ as the additive genetic variance divided by the total phenotypic variance [87]. Models were run using a parameter expanded prior ($V = 1$, nu $= 1$, alpha.mu $= 0$, alpha. $V = 1000$) for 500 000 iterations, with a burn in period of 50 000 and a thinning interval of 100 iterations [88].

## 3. Results

### 3.1. Overall performance

The 45 birds that completed the cylinder task in 2021 took between 7 and 24 trials (Mean ± s.d. = 10.0 ± 4.9) to pass habituation and training, and between 6 and 86 trials (Mean ± s.d. = 21.5 ± 20.2) to pass the test. In the 2022 U-barrier detour task, the 45 birds that completed the task took between 7 and 11 trials to pass habituation and training (Mean ± s.d. = 7.1 ± 0.6), and between 6 and 38 trials (Mean ± s.d. = 16.1 ± 8.9) to pass the test.

Birds who reached criterion in the associative learning task required between 10 and 94 trials to pass in 2021 ($N = 39$, Mean ± s.d. = 29.8 ± 19.5) and between 10 and 90 trials in 2022 ($N = 37$, Mean ± s.d. = 26.9 ± 16.5). Birds who reached criterion in the reversal learning task required between 13 and 114 trials in 2021 ($N = 35$, Mean ± s.d. = 50.4 ± 28.9), and between 12 and 94 trials in 2022 ($N = 33$, Mean ± s.d. = 42.1 ± 19.7). Wilcoxon signed rank tests showed that birds who reached criterion in both tasks required significantly more trials to do so during the reversal compared to the association task in both years (2021: $N = 35$, $Z = -3.645$, $p < 0.001$; 2022: $N = 33$, $Z = -3.664$, $p < 0.001$).

### 3.2. Relationships between tasks

There were no significant correlations between performances on the detour, association or reversal learning task in 2021 or 2022 (table 1). One component with an eigenvalue $> 1$ was extracted from the

**Table 1.** Results of Spearman correlation for task performances in 2021 ($N = 37$) and 2022 ($N = 35$).

| | | | reversal learning task | detour task |
|---|---|---|---|---|
| 2021 | detour task | $r_s$ | −0.064 | |
| | | $p$ | 0.708 | |
| | associative learning task | $r_s$ | 0.253 | −0.102 |
| | | $p$ | 0.130 | 0.547 |
| 2022 | detour task | $r_s$ | −0.295 | |
| | | $p$ | 0.085 | |
| | associative learning task | $r_s$ | 0.216 | −0.049 |
| | | $p$ | 0.214 | 0.782 |

PCA of task performances in both years, explaining 49.8% of the variation in task performance in 2021 and 47.4% in 2022. In both years, detour task performance loaded moderately and positively onto PC1, while association and reversal learning task performance loaded moderately and negatively (table 2).

## 3.3. Factors affecting task performance

During boldness tests, toutouwai showed low but significant repeatability in their latency to retrieve the mealworm whilst adjusting for testing year ($R \pm$ s.e. $= 0.281 \pm 0.084$, 95% CI $= 0.121$–$0.443$, $p < 0.001$). Body condition had no effect on latency to retrieve the mealworm during the 2022 boldness tests ($\beta \pm$ s.e. $= 2.031 \pm 4.474$, 95% CI $= -6.715$ to $10.865$, $p = 0.653$).

We found no effect of boldness, or any other non-cognitive factor we examined (e.g. age, sex, body condition), on performance in the detour (electronic supplementary material, tables S2 and S3), association (electronic supplementary material, table S4) or reversal learning tasks (electronic supplementary material, table S5). Moreover, prior experience did not improve performance: birds who participated in 2021 did not outperform naive birds during the associative learning task in 2022 ($Z = -0.032$, $p = 0.974$), reversal learning task ($Z = -1.008$, $p = 0.313$) or U-barrier detour task (electronic supplementary material, table s3). Support for the top three models of performance in each task is shown in electronic supplementary material, table S6.

## 3.4. Repeatability and heritability of task performance

Toutouwai showed low to moderate repeatability in their detour performance between years and barrier shape ($R \pm$ s.e. $= 0.360 \pm 0.153$, 95% CI $= 0.042$–$0.676$, $p = 0.024$) (figure 2a). They also showed moderate to high between-year repeatability in their performance in the association task ($R \pm$ s.e. $= 0.649 \pm 0.133$, 95% CI $= 0.342$–$0.804$, $p < 0.001$) (figure 2b) and moderate repeatability in the reversal learning task ($R \pm$ s.e. $= 0.511$, 95% CI $= 0.147$–$0.738$, $p = 0.009$) (figure 2c). However, there was no evidence for heritability of performance in the cylinder detour ($H^2 = 0.051$, 95% CI $= 0.000$–$0.186$), U-barrier detour ($H^2 = 0.084$, 95% CI $= 0.000$–$0.272$), association ($H^2 = 0.097$, 95% CI $= 0.000$–$0.360$) and reversal learning tasks ($H^2 = 0.105$, 95% CI $= 0.000$–$0.307$). 95% credible intervals indicated substantial uncertainty surrounding all estimates, and lower bounds did not differ from zero [88].

# 4. Discussion

We found no association between an individual's performance on detour and spatial reversal learning tasks. However, performance within each of these tasks was robust to potential confounds and repeatable across years. This suggests that detour and reversal learning tasks are reliable cognitive measures but are not underpinned by the same ability. Yet despite individual consistency in task performance, we found no evidence for heritability, potentially suggesting that inhibitory control is a plastic trait in toutouwai.
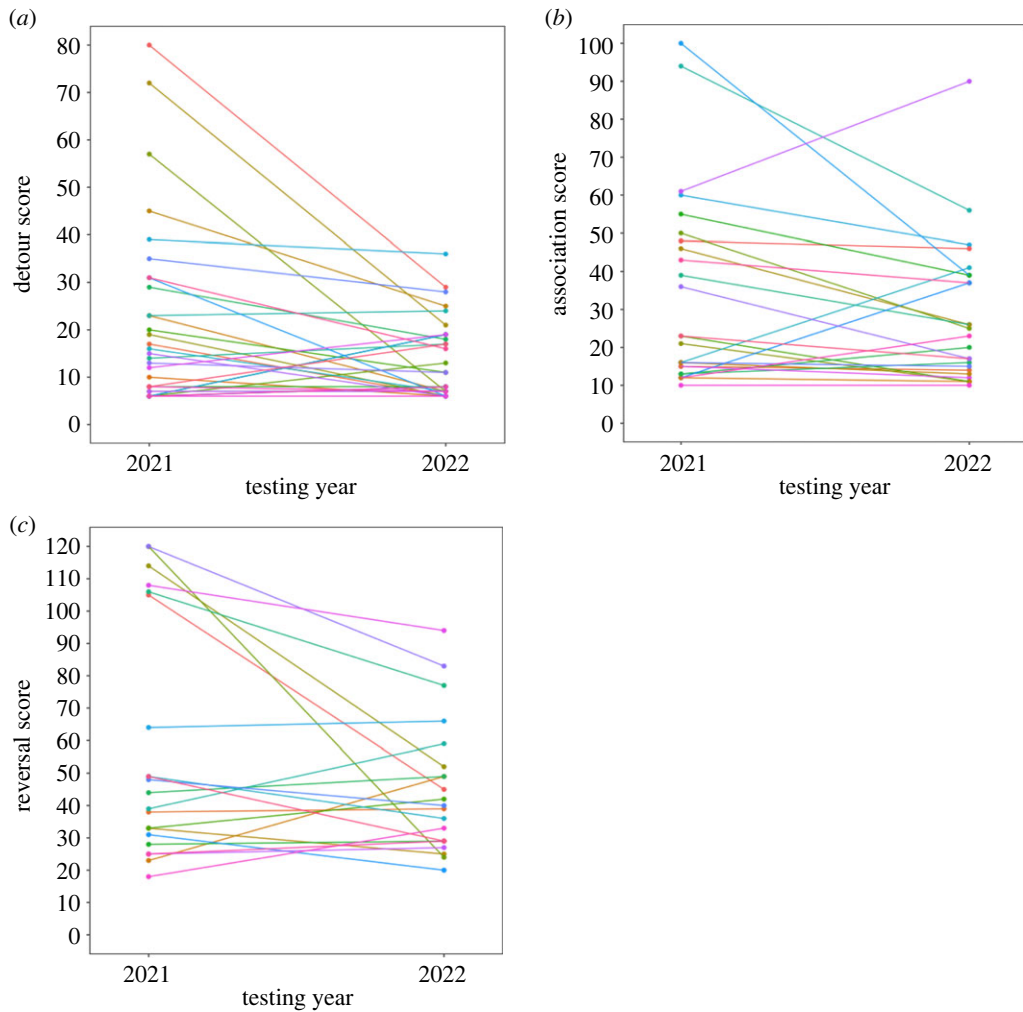
**Figure 2.** Relationship between individual performances in 2021 and 2022 on the (*a*) detour task, (*b*) associative learning task, and (*c*) reversal learning task. Circles represent individual performances with repeated measures connected by lines.

**Table 2.** Results of principal components analysis for task performances in 2021 ($N = 37$) and 2022 ($N = 35$).

| | 2021 | 2022 |
|---|---|---|
| task | PC1 | PC1 |
| detour task | 0.363 | 0.527 |
| associative learning task | −0.669 | −0.550 |
| reversal learning task | −0.649 | −0.648 |
| eigenvalue | 1.494 | 1.423 |
| % variance explained | 49.796 | 47.432 |

## 4.1. Relationships between tasks

We found no significant correlations between task performance measures, yet our PCA results suggested that the association and reversal learning tasks were at least partly underpinned by the same ability. This is expected, since both required individuals to form a spatial association using the same apparatus. By contrast, PCA results indicated that detour and reversal learning tasks captured distinct abilities. Both tasks evoked the expected prepotent responses and therefore seemed to require inhibitory control. Toutouwai readily learned the detour motor action when the barrier was opaque, but regularly pecked at the transparent barrier during the test phase [20]. Likewise, they required more trials to pass the reversal phase compared with the association, suggesting that they struggled to inhibit acting

on their knowledge of the previously rewarded location [7]. However, despite these superficial similarities, the lack of correlation between task performances suggests that the suppression of prepotent physical responses (motor inhibition) is a distinct cognitive ability from the suppression of prepotent mental representations (cognitive inhibition) in toutouwai. Alternatively, as the reversal learning task required cognitive flexibility and spatial associative learning in addition to inhibitory control [22,89], any shared reliance on inhibitory control during the detour and reversal tasks may have been too small to detect during our study. Additionally, as tasks were conducted at different times of year, it is possible that temporal behavioural variation affected performance and obscured correlations between tasks. However, toutouwai do not show any major behavioural differences across the non-breeding season [59,90], which is when testing took place. Unmeasured confounding variables, such as other aspects of personality, may have also affected each task differently and masked positive relationships between them [25]. However, the literature overwhelmingly supports a lack of convergent validity across a range of inhibitory control tasks in non-human animal species, including dogs [2,26,28], birds [29,91], primates [25] and lizards [27]. Therefore, it seems likely that our data suggest that inhibitory control is not a unitary construct in toutouwai, with motor and cognitive inhibition functioning as separable components of this ability.

## 4.2. Factors affecting task performance

An ongoing concern in inhibitory control research is the extent to which tasks capture non-cognitive variables [3,4]. Reassuringly, we found no effect of boldness, body condition or prior experience on any task performance. Likewise, we did not find an effect of sex or age. Two previous studies have assessed the influence of body condition on cylinder detour performance in toutouwai, using very similar tasks but producing conflicting results. Shaw et al. [60] found no correlation between condition and performance, but a later study with a larger sample size found that birds in better condition required fewer trials to pass [4]. Interestingly, the average number of trials toutouwai required to reach criterion in [4] was roughly half that of Shaw et al. [60] and the current study. This suggests that subtle differences in the environment or experimental procedures may have affected how toutouwai performed in [4], leading to better performance and generating a body condition effect. These inconsistent findings highlight a more general problem with replicability in the cognitive sciences [92], one that is particularly relevant to field experiments where standardization is exceedingly difficult [93].

Birds who participated in the cylinder detour task did not perform significantly better than naive birds on the U-barrier detour the following year, a finding corroborated by previous studies of toutouwai [4] and guppies [94]. While prior experience improved performance on a novel detour task three days later in pheasants [3], toutouwai do not appear to remember the properties of transparent barriers after 1 year, or are unable to generalize their knowledge to a visually distinct apparatus requiring a different motor action [20]. Similarly, although animals can improve their performance over multiple reversals conducted in quick succession [95,96] potentially by adopting a win-stay/lose-shift strategy and anticipating upcoming reversals [22], experience on the association/reversal learning tasks did not improve task performance after a 1-year interval in toutouwai.

Toutouwai showed low but significant repeatability in their latencies to take a mealworm under threat. Although we did not measure contextual repeatability and thus cannot be sure that we captured a personality trait, this temporal consistency does provide some evidence that our test measured boldness [63]. Furthermore, body condition did not affect latency to retrieve the mealworm, suggesting that this test was not simply measuring hunger. Despite boldness being a predictor of task performance in other species [32–34], we found no evidence for this in toutouwai. We anticipated that shyer individuals may be more fearful during testing, which can impair inhibitory control [36]. However, our habituation procedures may have ensured that the apparatus and experimenter were no longer perceived as risky during testing [17]. Equally, there was little indication that personality led to selection bias in experiments. In each year, we tested the majority of toutouwai in the study population, and in only one case was a bird unwilling (i.e. potentially too shy/neophobic) to interact with the apparatus.

## 4.3. Repeatability and heritability of task performance

In all tasks, toutouwai were repeatable in their performance across a 1-year interval. Consistency was highest for the associative learning task ($R = 0.649$), followed by the reversal learning task ($R = 0.511$) and the detour task ($R = 0.360$). All values exceeded the average temporal repeatability of cognitive performance found in a meta-analysis by Cauchoix et al. [38] ($R = 0.18–0.28$). The lower repeatability

observed in the detour task may be because the tasks were visually distinct and required different locomotor actions. Equally, our results may indicate that associative learning ability (or more specifically, spatial associative learning) is more temporally stable than inhibitory control, supported by the fact that reversal learning (putatively capturing both associative learning and inhibition) was less repeatable than associative learning alone. Similar repeatability results were found in wild Australian magpies, where repeatability over a 3-year period was strongest for associative learning, somewhat lower for reversal learning, and non-significant for a detour task [41]. However, it should also be noted that repeatability values for task performance may not necessarily reflect the true repeatability of cognitive traits [38]. We cannot exclude the possibility that our estimates also captured consistency (or lack thereof) in non-cognitive confounds of tasks performance that were unmeasured in our study [97]. Nonetheless, our results do provide some evidence that these tasks captured consistent individual differences in cognition.

Surprisingly, Shaw [4] found no repeatability over a 1-year interval for cylinder detour tasks in toutouwai. However, the smaller sample size used in this previous study (15 individuals tested twice, compared to 31 in the current study) may have reduced the reliability of parameter estimates [79]. Toutouwai in this previous study also performed substantially better with less inter-individual variation during their second task presentation compared to the current study. This may have affected repeatability estimates, as these measure the amount of inter-individual variation relative to intra-individual variation [39].

Despite finding consistent individual variation in all tasks, we found no evidence that task performance was heritable. Heritability estimates were low and non-significant for all tasks, ranging from 0.051 to 0.105. However, the precision of these estimates was poor: 95% credible intervals indicated that true heritabilities could range from negligible to moderate [88]. This may be a consequence of our small sample size and missing pedigree information for some subjects, both of which affect the power to detect heritability using animal models [98]. Our results are in line with studies of Aegean wall lizards (*Podarcis erhardii*) and great tits, which found no evidence for heritability in reversal learning task performance [49,50]. By contrast, reversal learning performance showed moderate heritability in red junglefowl (*Gallus gallus*) [51] and house mice (*Mus musculus*) [53], as did detour task performance in pheasants [52]. Since repeatability was high compared with heritability, inhibitory control variation may be largely shaped by environmental factors in toutouwai, the effects of which remained relatively stable over the 1 year period that repeatability was assessed [39,43].

Plasticity of inhibitory control performance in non-human animals has been reported by other studies. Aegean wall lizards from less variable physical environments were better at reversal learning [99] but this trait shows no heritability in this species [49]. Conversely, pheasants reared in spatially unpredictable environments showed better detour performance [100]. Social factors can also shape inhibitory control development. Group size was positively related to detouring ability in spotted hyaenas (*Crocuta crocuta*) [101], and Australian magpies from larger groups performed better on a suite of cognitive tasks, including detour and reversal learning [5]. Toutouwai are largely solitary [59], so social factors may be comparatively unimportant for this species. However, characteristics of the physical environment could explain the stable cognitive differences we observed, as toutouwai often retain the same territory for life [102].

Obtaining more robust heritability estimates and exploring non-genetic influences on inhibitory control will be a valuable avenue for future research in toutouwai. Additionally, exploring relationships between inhibitory control and fitness would be helpful in interpreting our heritability estimates, since traits important to fitness often show low heritability due to strong directional selection depleting genetic variation [39,103]. Toutouwai of both sexes cache to manage short-term fluctuations in food availability [104], and males share food with their mates during courtship and incubation [59]. Both behaviours likely require birds to inhibit eating food immediately, so inhibitory control may predict survival and reproductive success in this species when selection favours caching and food-sharing over immediately eating food as a behavioural tactic. However, our low heritability estimates for all tasks (albeit preliminary) suggest that the cognitive traits underpinning performance would evolve relatively slowly in toutouwai, even if they were subject to strong selection [43].

# 5. Conclusion

Despite increasing interest in the evolution of inhibitory control, little is known about the long-term repeatability and heritability of this ability in the wild or the validity of widespread testing

paradigms. In wild toutouwai, we found no evidence for a 'general' inhibitory control ability underpinning performance across detour and reversal learning tasks, supporting the growing consensus that inhibitory control is best conceptualized as an umbrella term capturing distinct abilities in many species. Multi-task approaches will be necessary to understand the evolution of inhibitory control in species where this is not a unitary construct, since we cannot assume that all components have the same genetic basis or relationship to fitness. Despite this, detour and reversal learning tasks appear to be valuable measures of motor and cognitive inhibition, respectively, in toutouwai, as performance was repeatable across years and unaffected by potential confounds. However, we found no evidence that task performance was heritable, suggesting that individual differences in inhibitory control may be plastic in this species. Our study highlights the importance of dissecting what cognitive tasks measure, as this determines the conclusions we can draw about the causes and consequences of cognitive variation.

# References

1. Boogert NJ, Anderson RC, Peters S, Searcy WA, Nowicki S. 2011 Song repertoire size in male song sparrows correlates with detour reaching, but not with other cognitive measures. *Anim. Behav.* **81**, 1209–1216. (doi:10.1016/j.anbehav.2011.03.004)

2. Brucks D, Marshall-Pescini S, Wallis LJ, Huber L, Range F. 2017 Measures of dogs' inhibitory control abilities do not correlate across tasks. *Front. Psychol.* **8**, 849–849. (doi:10.3389/fpsyg.2017.00849)

3. van Horik JO, Langley EJG, Whiteside MA, Laker PR, Beardsworth CE, Madden JR. 2018 Do detour tasks provide accurate assays of inhibitory control? *Proc. R. Soc. B* **285**, 20180150. (doi:10.1098/rspb.2018.0150)

4. Shaw RC. 2017 Testing cognition in the wild: factors affecting performance and individual consistency in two measures of avian cognition. *Behav. Processes* **134**, 31–36. (doi:10.1016/j.beproc.2016.06.004)

5. Ashton BJ, Ridley AR, Edwards EK, Thornton A. 2018 Cognitive performance is linked to group size and affects fitness in Australian magpies. *Nature* **554**, 364–367. (doi:10.1038/nature25503)

6. Davidson GL, Reichert MS, Coomes JR, Kulahci IG, de la Hera I, Quinn JL. 2022 Inhibitory control performance is repeatable over time and across contexts in a wild bird population. *Anim. Behav.* **187**, 305–318. (doi:10.1016/j.anbehav.2022.02.007)

7. Diamond A. 2013 Executive functions. *Annu. Rev. Psychol.* **64**, 135–168. (doi:10.1146/annurev-psych-113011-143750)

8. Tiego J, Testa R, Bellgrove MA, Pantelis C, Whittle S. 2018 A hierarchical model of inhibitory control. *Front. Psychol.* **9**, 1339. (doi:10.3389/fpsyg.2018.01339)

9. Moffitt T et al. 2011 A gradient of childhood self-control predicts health, wealth, and public safety. *Proc. Natl Acad. Sci. USA* **108**, 2693–2698. (doi:10.1073/pnas.1010076108)

10. Miller R, Boeckle M, Jelbert SA, Frohnwieser A, Wascher CAF, Clayton NS. 2019 Self-control in crows, parrots and nonhuman primates. *WIREs Cogn. Sci.* **10**, e1504. (doi:10.1002/wcs.1504)

11. Stevens JR, Stephens D. 2008 Patience. *Curr. Biol.* **18**, 11–12. (doi:10.1016/j.cub.2007.11.021)

12. Stevens JR, Kennedy BA, Morales D, Burks M. 2016 The domain specificity of intertemporal choice in pinyon jays. *Psychon. Bull. Rev.* **23**, 915–921. (doi:10.3758/s13423-015-0973-6)

13. MacLean EL et al. 2014 The evolution of self-control. *Proc. Natl Acad. Sci. USA* **111**, 2140–2148. (doi:10.1073/pnas.1323533111)

14. Madden JR, Langley EJG, Whiteside MA, Beardsworth CE, van Horik JO. 2018 The quick are the dead: pheasants that are slow to reverse a learned association survive for longer in the wild. *Phil. Trans. R. Soc. B* **373**, 20170297. (doi:10.1098/rstb.2017.0297)

15. Isden J, Panayi C, Dingle C, Madden J. 2013 Performance in cognitive and problem-solving tasks in male spotted bowerbirds does not correlate with mating success. *Anim. Behav.* **86**, 829–838. (doi:10.1016/j.anbehav.2013.07.024)

16. MacKinlay RD, Shaw RC. 2019 Male New Zealand robin (*Petroica longipes*) song repertoire size does not correlate with cognitive performance in the wild. *Intelligence* **74**, 25–33. (doi:10.1016/j.intell.2018.10.009)

17. Morand-Ferron J, Cole EF, Quinn JL. 2016 Studying the evolutionary ecology of cognition in the wild: a review of practical and conceptual challenges. *Biol. Rev.* **91**, 367–389. (doi:10.1111/brv.12174)

18. Thornton A, Lukas D. 2012 Individual variation in cognitive performance: developmental and evolutionary perspectives. *Phil. Trans. R. Soc. B* **367**, 2773–2783. (doi:10.1098/rstb.2012.0214)

19. Nigg JT. 2017 Annual Research Review: On the relations among self-regulation, self-control, executive functioning, effortful control, cognitive control, impulsivity, risk-taking, and inhibition for developmental psychopathology. *J. Child Psychol. Psychiatry* **58**, 361–383. (doi:10.1111/jcpp.12675)

20. Kabadayi C, Bobrowicz K, Osvath M. 2018 The detour paradigm in animal cognition. *Anim. Cogn.* **21**, 21–35. (doi:10.1007/s10071-017-1152-0)

21. Anderson MC, Green C. 2001 Suppressing unwanted memories by executive control. *Nature* **410**, 366–369. (doi:10.1038/35066572)

22. Izquierdo A, Brigman JL, Radke AK, Rudebeck PH, Holmes A. 2017 The neural basis of reversal learning: an updated perspective. *Neuroscience* **345**, 12–26. (doi:10.1016/j.neuroscience.2016.03.021)

23. Diamond A, Ling DS. 2016 Conclusions about interventions, programs, and approaches for improving executive functions that appear justified and those that, despite much hype, do not. *Dev. Cogn. Neurosci.* **18**, 34–48. (doi:10.1016/j.dcn.2015.11.005)

24. Cronbach LJ, Meehl PE. 1955 Construct validity in psychological tests. *Psychol. Bull.* **52**, 281–302. (doi:10.1037/h0040957)

25. Völter CJ, Tinklenberg B, Call J, Seed AM. 2018 Comparative psychometrics: establishing what differs is central to understanding what evolves. *Phil. Trans. R. Soc. B* **373**, 20170283. (doi:10.1098/rstb.2017.0283)

26. Bray EE, MacLean EL, Hare BA. 2014 Context specificity of inhibitory control in dogs. *Anim. Cogn.* **17**, 15–31. (doi:10.1007/s10071-013-0633-z)

27. Szabo B, Hoefer S, Whiting MJ. 2020 Are lizards capable of inhibitory control? Performance on a semi-transparent version of the cylinder task in five species of Australian skinks. *Behav. Ecol. Sociobiol.* **74**, 115–123. (doi:10.1007/s00265-020-02897-y)

28. Vernouillet AAA, Stiles LR, Andrew McCausland J, Kelly DM. 2018 Individual performance across motoric self-regulation tasks are not correlated for pet dogs. *Learn. Behav.* **46**, 522–536. (doi:10.3758/s13420-018-0354-x)

29. Troisi CA, Cooke AC, Davidson GL, Hera I, Reichert MS, Quinn JL. 2021 No evidence for cross-contextual consistency in spatial learning and behavioural flexibility in a passerine. *Anim. Behav. Cogn.* **8**, 46–461. (doi:10.26451/abc.08.03.08.2021)

30. Réale D, Dingemanse NJ, Kazem AJN, Wright J. 2010 Evolutionary and ecological approaches to the study of personality. *Phil. Trans. R. Soc. B* **365**, 3937–3946. (doi:10.1098/rstb.2010.0222)

31. Dougherty LR, Guillette LM. 2018 Linking personality and cognition: a meta-analysis. *Phil. Trans. R. Soc. B* **373**, 20170282. (doi:10.1098/rstb.2017.0282)

32. Gomes ACR, Guerra S, Silva PA, Marques C, Trigo S, Boogert NJ, Cardoso GC. 2020 Proactive common waxbills make fewer mistakes in a cognitive assay, the detour-reaching task. *Behav. Ecol. Sociobiol.* **74**, 31. (doi:10.1007/s00265-020-2809-2)

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

34. Bensky MK, Bell AM. 2020 Predictors of individual variation in reversal learning performance in three-spined sticklebacks. *Anim. Cogn.* **23**, 925–938. (doi:10.1007/s10071-020-01399-8)

35. Shaw RC, Schmelz M. 2017 Cognitive test batteries in animal cognition research: evaluating the past, present and future of comparative psychometrics. *Anim. Cogn.* **20**, 1003–1018. (doi:10.1007/s10071-017-1135-1)

36. Hartikainen KM, Siiskonen AR, Ogawa KH. 2012 Threat interferes with response inhibition. *Neuroreport* **23**, 447–450. (doi:10.1097/WNR.0b013e3283531e74)

37. Thornton A, Isden J, Madden JR. 2014 Toward wild psychometrics: linking individual cognitive differences to fitness. *Behav. Ecol.* **25**, 1299–1301. (doi:10.1093/beheco/aru095)

38. Cauchoix M et al. 2018 The repeatability of cognitive performance: a meta-analysis. *Phil. Trans. R. Soc. B* **373**, 20170281. (doi:10.1098/rstb.2017.0281)

39. Falconer DS, Mackay TFC. 1996 *Introduction to quantitative genetics*, 4th edn. Essex: Longman Press.

40. Bell AM, Hankison SJ, Laskowski KL. 2009 The repeatability of behaviour: a meta-analysis. *Anim. Behav.* **77**, 771–783. (doi:10.1016/j.anbehav.2008.12.022)

41. Ashton BJ, Thornton A, Cauchoix M, Ridley AR. 2022 Long-term repeatability of cognitive performance. *R. Soc. Open Sci.* **9**, 220069. (doi:10.1098/rsos.220069)

42. Griffin AS, Guillette LM, Healy SD. 2015 Cognition and personality: an analysis of an emerging field. *Trends Ecol. Evol.* **30**, 207–214. (doi:10.1016/j.tree.2015.01.012)

43. Boake CRB. 1989 Repeatability: Its role in evolutionary studies of mating behavior. *Evol. Ecol.* **3**, 173–182. (doi:10.1007/BF02270919)

44. Visscher PM, Hill WG, Wray NR. 2008 Heritability in the genomics era - concepts and misconceptions. *Nat. Rev. Genet.* **9**, 255–266. (doi:10.1038/nrg2322)

45. Willems YE, Boesen N, Li J, Finkenauer C, Bartels M. 2019 The heritability of self-control: a meta-analysis. *Neurosci. Biobehav. Rev.* **100**, 324–334. (doi:10.1016/j.neubiorev.2019.02.012)

46. Schachar RJ, Forget-Dubois N, Dionne G, Boivin M, Robaey P. 2011 Heritability of response inhibition in children. *J. Int. Neuropsychol. Soc.* **17**, 238–247. (doi:10.1017/S1355617710001463)

47. Friedman NP, Miyake A, Young SE, DeFries JC, Corley RP, Hewitt JK. 2008 Individual differences in executive functions are almost entirely genetic in origin. *J. Exp. Psychol.-Gen.* **137**, 201–225. (doi:10.1037/0096-3445.137.2.201)

48. Crosbie J et al. 2013 Response inhibition and ADHD traits: correlates and heritability in a community sample. *J. Abnorm. Child Psychol.* **41**, 497–507. (doi:10.1007/s10802-012-9693-9)

49. De Meester G, Pafilis P, Vasilakis G, Van Damme R. 2022 Exploration and spatial cognition show long-term repeatability but no heritability in the Aegean wall lizard. *Anim. Behav.* **190**, 167–185. (doi:10.1016/j.anbehav.2022.06.007)

50. van den Heuvel K, Quinn JL, Kotrschal A, van Oers K. 2023 Artificial selection for reversal learning reveals limited repeatability and no heritability of cognitive flexibility in great tits (*Parus major*). *Proc. R. Soc. B* **290**, 20231067. (doi:10.1098/rspb.2023.1067)

51. Sorato E, Zidar J, Garnham L, Wilson A, Løvlie H. 2018 Heritabilities and co-variation among cognitive traits in red junglefowl. *Phil. Trans. R. Soc. Lond. B* **373**, 20170285. (doi:10.1098/rstb.2017.0285)

52. Langley EJG, Adams G, Beardsworth CE, Dawson DA, Laker PR, van Horik JO, Whiteside MA, Wilson AJ, Madden JR. 2020 Heritability and correlations among learning and inhibitory control traits. *Behav. Ecol.* **31**, 798–806. (doi:10.1093/beheco/araa029)

53. Laughlin RE, Grant TL, Williams RW, Jentsch JD. 2011 Genetic dissection of behavioral flexibility: reversal learning in mice. *Biol. Psychiatry* **69**, 1109–1116. (doi:10.1016/j.biopsych.2011.01.014)

54. Croston R, Branch CL, Kozlovsky DY, Dukas R, Pravosudov VV. 2015 Heritability and the evolution of cognitive traits. *Behav. Ecol.* **26**, 1447–1459. (doi:10.1093/beheco/arv088)

55. Sauce B, Bendrath S, Herzfeld M, Siegel D, Style C, Rab S, Korabelnikov J, Matzel LD. 2018 The impact of environmental interventions among mouse siblings on the heritability and malleability of general cognitive ability. *Phil. Trans. R. Soc. B* **373**, 20170289. (doi:10.1098/rstb.2017.0289)

56. Bell G. 2007 Artificial selection. In *Selection: The mechanism of evolution* (ed. G Bell), pp. 186–221. Oxford, UK: Oxford University Press.

57. Whitlock MC, Fowler K. 1999 The changes in genetic and environmental variance with inbreeding in *Drosophila melanogaster*. *Genetics* **152**, 345–353.

58. Howrigan DP et al. 2016 Genome-wide autozygosity is associated with lower general cognitive ability. *Mol. Psychiatry* **21**, 837–843. (doi:10.1038/mp.2015.120)

59. Higgins PJ, Peter JM. 2002 New Zealand robin. In *Handbook of Australian, New Zealand and antarctic birds*, pp. 706–725. Melbourne, Australia: Oxford University Press.

60. Shaw RC, Boogert NJ, Clayton NS, Burns KC. 2015 Wild psychometrics: evidence for 'general' cognitive performance in wild New Zealand robins, *Petroica longipes*. *Anim. Behav.* **109**, 101–111. (doi:10.1016/j.anbehav.2015.08.001)

61. Carter AJ, Feeney WE, Marshall HH, Cowlishaw G, Heinsohn R. 2013 Animal personality: what are behavioural ecologists measuring? *Biol. Rev.* **88**, 465–475. (doi:10.1111/brv.12007)

62. Hardt O, Nader K, Nadel L. 2013 Decay happens: the role of active forgetting in memory. *Trends Cogn. Sci.* **17**, 111–120. (doi:10.1016/j.tics.2013.01.001)

63. Réale D, Reader SM, Sol D, McDougall PT, Dingemanse NJ. 2007 Integrating animal temperament within ecology and evolution. *Biol. Rev.* **82**, 291–318. (doi:10.1111/j.1469-185X.2007.00010.x)

64. Friard O, Gamba M. 2016 BORIS: a free, versatile open-source event-logging software for video/audio coding and live observations. *Methods Ecol. Evol.* **7**, 1325–1330.

65. Cauchard L, Boogert NJ, Lefebvre L, Dubois F, Doligez B. 2013 Problem-solving performance is correlated with reproductive success in a wild bird population. *Anim. Behav.* **85**, 19–26. (doi:10.1016/j.anbehav.2012.10.005)

66. Armstrong D. 2001 Sexing North Island robins (*Petroica australis longipes*) from morphometrics and plumage. *Notornis* **48**, 76–80.

67. Powlesland RG. 1983 Seasonal and diurnal variation in vocal behaviour of the South Island robin. *N. Z. J. Zool.* **10**, 225–232. (doi:10.1080/03014223.1983.10423908)

68. R Core Team. 2022 R: A language and environment for statistical computing.

69. Brooks ME, Kristensen K, van Benthem KJ, Magnusson A, Berg CW, Nielsen A, Skaug HJ, Mächler M, Bolker BM. 2017 glmmTMB balances speed and flexibility among packages

for zero-inflated generalized linear mixed modelling. *R J.* **9**, 378–400.

70. Bartoń K. 2020 MuMIn: multi-model inference.

71. Stoffel MA, Nakagawa S, Schielzeth H. 2017 rptR: repeatability estimation and variance decomposition by generalized linear mixed-effects models. *Methods Ecol. Evol.* **8**, 1639–1644. (doi:10.1111/2041-210X.12797)

72. Hadfield JD. 2010 MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R Package. *J. Stat. Softw.* **33**, 1–22.

73. Harrell FE. 2023 Hmisc: Harrell miscellaneous.

74. Kassambara A, Mundt F. 2020 factoextra: Extract and visualize the results of multivariate data analyses.

75. Harper DGC. 1994 Some comments on the repeatability of measurements. *Ringing Migr.* **15**, 84–90. (doi:10.1080/03078698.1994.9674078)

76. Guttman L. 1954 Some necessary conditions for common-factor analysis. *Psychometrika* **19**, 149–161. (doi:10.1007/BF02289162)

77. Kaiser HF. 1960 The application of electronic computers to factor analysis. *Educ. Psychol. Meas.* **20**, 141–151. (doi:10.1177/001316446002000116)

78. Burnham KP, Anderson DR. 2002 *Model selection and multimodel inference: a practical information-theoretic approach*, 2nd edn. New York: NY: Springer.

79. Nakagawa S, Schielzeth H. 2010 Repeatability for Gaussian and non-Gaussian data: a practical guide for biologists. *Biol. Rev.* **85**, 935–956. (doi:10.1111/j.1469-185X.2010.00141.x)

80. Martin JGA, Nussey DH, Wilson AJ, Réale D. 2011 Measuring individual differences in reaction norms in field and experimental studies: a power analysis of random regression models. *Methods Ecol. Evol.* **2**, 362–374. (doi:10.1111/j.2041-210X.2010.00084.x)

81. Ardern SL, Ma W, Ewen JG, Armstrong DP, Lambert DM. 1997 Social and sexual monogamy in translocated New Zealand robin populations detected using minisatellite DNA. *The Auk* **114**, 120–126. (doi:10.2307/4089074)

82. de Villemereuil P, Gimenez O, Doligez B. 2013 Comparing parent–offspring regression with frequentist and Bayesian animal models to estimate heritability in wild populations: a simulation study for Gaussian and binary traits. *Methods Ecol. Evol.* **4**, 260–275. (doi:10.1111/2041-210X.12011)

83. Kruuk LEB, Hadfield JD. 2007 How to separate genetic and environmental causes of similarity between relatives. *J. Evol. Biol.* **20**, 1890–1903. (doi:10.1111/j.1420-9101.2007.01377.x)

84. Cameron TC, Plaistow S, Mugabo M, Piertney SB, Benton TG. 2014 Eco-evolutionary dynamics: Experiments in a model system. In *Advances in ecological research* (eds J Moya-Laraño, J Rowntree, G Woodward), pp. 171–206. New York, NY: Academic Press.

85. Wolak ME. 2012 nadiv: an R package to create relatedness matrices for estimating non-additive genetic variances in animal models. *Methods Ecol. Evol.* **3**, 792–796. (doi:10.1111/j.2041-210X.2012.00213.x)

86. Thomson CE, Winney IS, Salles OC, Pujol B. 2018 A guide to using a multiple-matrix animal model to disentangle genetic and nongenetic causes of phenotypic variance. *PLoS ONE* **13**, e0197720. (doi:10.1371/journal.pone.0197720)

87. de Villemereuil P, Schielzeth H, Nakagawa S, Morrissey M. 2016 General methods for evolutionary quantitative genetic inference from generalized mixed models. *Genetics* **204**, 1281–1294. (doi:10.1534/genetics.115.186536)

88. de Villemereuil P. 2021 Estimation of a biological trait heritability using the animal model and MCMCglmm. See https://devillemereuil.legtux.org/wp-content/uploads/2021/09/tuto_en.pdf (accessed on 9 October 2022).

89. Izquierdo A, Jentsch JD. 2012 Reversal learning as a measure of impulsive and compulsive behavior in addictions. *Psychopharmacology (Berl.)* **219**, 607–620. (doi:10.1007/s00213-011-2579-7)

90. Steer J, Burns K. 2008 Seasonal variation in male-female competition, cooperation and selfish hoarding in a monogamous songbird. *Behav. Ecol. Sociobiol.* **62**, 1175–1183. (doi:10.1007/s00265-008-0546-z)

91. Anderson RC, Searcy WA, Peters S, Hughes M, DuBois AL, Nowicki S. 2017 Song learning and cognitive ability are not consistently related in a songbird. *Anim. Cogn.* **20**, 309–320. (doi:10.1007/s10071-016-1053-7)

92. Farrar BG, Boeckle M, Clayton NS. 2020 Replications in comparative cognition: What should we expect and how can we improve? *Anim. Behav. Cogn.* **7**, 1–22. (doi:10.26451/abc.07.01.02.2020)

93. Filazzola A, Cahill Jr JF. 2021 Replication in field ecology: identifying challenges and proposing solutions. *Methods Ecol. Evol.* **12**, 1780–1792. (doi:10.1111/2041-210X.13657)

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

95. Bond AB, Kamil AC, Balda RP. 2007 Serial reversal learning and the evolution of behavioral flexibility in three species of North American corvids (*Gymnorhinus cyanocephalus, Nucifraga columbiana,* Aphelocoma californica). *J. Comp. Psychol.* **121**, 372–379. (doi:10.1037/0735-7036.121.4.372)

96. Cauchoix M, Hermer E, Chaine AS, Morand-Ferron J. 2017 Cognition in the field: comparison of reversal learning performance in captive and wild passerines. *Sci. Rep.* **7**, 12945. (doi:10.1038/s41598-017-13179-5)

97. Wilson AJ. 2018 How should we interpret estimates of individual repeatability? *Evol. Lett.* **2**, 4–8. (doi:10.1002/evl3.40)

98. Wilson AJ, Réale D, Clements MN, Morrissey MM, Postma E, Walling CA, Kruuk LEB, Nussey DH. 2010 An ecologist's guide to the animal model. *J. Anim. Ecol.* **79**, 13–26. (doi:10.1111/j.1365-2656.2009.01639.x)

99. De Meester G, Sfendouraki-Basakarou A, Pafilis P, Damme RV. 2021 Dealing with the unexpected: the effect of environmental variability on behavioural flexibility in a Mediterranean lizard. *Behaviour* **158**, 1193–1223. (doi:10.1163/1568539X-bja10088)

100. van Horik JO, Beardsworth CE, Laker PR, Langley EJG, Whiteside MA, Madden JR. 2019 Unpredictable environments enhance inhibitory control in pheasants. *Anim. Cogn.* **22**, 1105–1114. (doi:10.1007/s10071-019-01302-0)

101. Johnson-Ulrich L, Holekamp KE. 2020 Group size and social rank predict inhibitory control in spotted hyaenas. *Anim. Behav.* **160**, 157–168. (doi:10.1016/j.anbehav.2019.11.020)

102. McGavin S. 2009 Density and pair fidelity in a translocated population of North Island robin (*Petroica longipes*). *Notornis* **56**, 206–212.

103. Mousseau TA, Roff DA. 1987 Natural selection and the heritability of fitness components. *Heredity* **59**, 181–197. (doi:10.1038/hdy.1987.113)

104. Menzies IJ, Burns KC. 2008 Food hoarding in the New Zealand robin: A review and synthesis. In *Animal behavior: new research* (eds EA Weber, LH Krause), pp. 163–183. New York, NY: Nova Science Publishers.

105. McCallum E, Shaw RC. 2023 Repeatability and heritability of inhibitory control performance in wild toutouwai (*Petroica longipes*). Figshare. (doi:10.6084/m9.figshare.c.6922247)