

Published in final edited form as:

Nature. 2018 February 15; 554(7692): 364–367. doi:10.1038/nature25503.

Cognitive performance is linked to group size and affects fitness in Australian magpies

Benjamin J. Ashton^{1,*}, Amanda R. Ridley^{1,*}, Emily K. Edwards¹, and Alex Thornton^{2,*}

¹Centre for Evolutionary Biology, School of Biological Sciences, University of Western Australia, 35 Stirling Highway, Crawley 6009, Australia

²Centre for Ecology and Conservation, University of Exeter, Penryn Campus, Treliever Road, Penryn TR10 9FE, UK

Abstract

The Social Intelligence Hypothesis argues that the demands of social life drive cognitive evolution^{1–3}. This idea receives support from comparative studies linking variation in group size or mating systems with cognitive and neuroanatomical differences across species^{3–7}, but findings are contradictory and contentious^{8–10}. To understand the cognitive consequences of sociality it is also important to investigate social variation *within* species. Here we show that in wild, cooperatively breeding Australian magpies, individuals living in larger groups show elevated cognitive performance, which in turn is linked to increased reproductive success. Individual performance was highly correlated across four cognitive tasks, hinting towards a “general intelligence factor” underlying cognitive performance. Repeated cognitive testing of juveniles at different ages showed that the group size – cognition correlation emerged in early life, suggesting that living in larger groups promotes cognitive development. Furthermore, we found a positive association between female task performance and three indicators of reproductive success, thus identifying a selective benefit of greater cognitive performance. Together, these results provide critical *intraspecific* evidence that sociality can shape cognitive development and evolution.

The social environment is commonly assumed to generate important cognitive challenges. According to the Social Intelligence (or Social Brain) Hypothesis, these challenges, including the need to form and maintain social bonds, track third party relationships and

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:http://www.nature.com/authors/editorial_policies/license.html#terms

* Correspondence and requests for materials should be addressed to B.J.A benjamin.ashton@research.uwa.edu.au; amanda.ridley@uwa.edu.au; alex.thornton@exeter.ac.uk.

Author contributions

B.J.A., A.R.R. and A.T. conceived and designed the study. B.J.A. wrote the manuscript. B.J.A. and E.K.E. carried out data collection. All authors discussed results and commented on the manuscript.

Author information

The data that support the findings of this study have been deposited in the Dryad Digital Repository (doi:[10.5061/dryad.ph3h8](https://doi.org/10.5061/dryad.ph3h8)). Reprints and permissions information is available at www.nature.com/reprints.

The authors declare no competing financial interests.

Data availability

The data that support the findings of this study have been deposited in the Dryad Digital Repository (doi:[10.5061/dryad.ph3h8](https://doi.org/10.5061/dryad.ph3h8)). Source data has been provided for Figures 1–4, and Extended Data Figures 2 and 3.

anticipate others' actions, are the central drivers of cognitive evolution^{1–3}. This argument receives widespread support from studies linking variation in social factors, such as group size or mating systems, with differences in cognitive performance or neuroanatomy across species of birds and mammals (e.g.^{3–6}). However, comparative analyses are subject to ecological and phylogenetic confounding effects, and have yielded conflicting results, with recent work calling into question the importance of social factors^{8–10}. To understand the role of sociality in cognitive evolution, it is critical to examine the causes and fitness consequences of cognitive variation *within* species^{11,12}.

In species living in stable social groups, within-population variation in group size could generate differences in information-processing demands and so influence the expression of cognitive traits¹³. Measures of brain structure correlate with group size in humans, captive cichlids (*Neolamprologus pulcher*) and captive macaques (*Macaca mulatta*)^{13–15}, but the relationship between group size and cognition in wild animals is unknown. Furthermore, the potential for group-size dependent cognitive traits to come under selection is not understood, as their fitness consequences have never been investigated. To address these critical gaps in our knowledge, we examined whether group size predicts individual variation in cognitive performance (controlling for morphological, nutritional and behavioural factors) within a population of wild, cooperatively breeding Australian magpies (Western Australian subspecies, *Cracticus tibicen dorsalis*). We quantified individual cognitive performance in 56 birds from 14 groups, ranging in size from 3–12 individuals, using a battery of cognitive tasks designed to measure inhibitory control, associative learning, reversal learning, and spatial memory (Extended Data Fig. 1). These four domain-general cognitive processes are thought to play an important role in a range of fitness-related behaviours in both social and asocial contexts^{11,16} (see supplementary methods for details).

Group size was the strongest predictor of adult performance across all four tasks (Tables S1 to S4), with individuals from larger groups outperforming those from smaller groups (Fig. 1). Individual performance was significantly positively correlated across all four tasks (Table S5) suggestive of an underlying “general intelligence” factor akin to that reported in human psychometric studies¹⁷. A principal components analysis (PCA) revealed performance in all four tasks loaded positively onto the first principal component (PC1; eigenvalue >1). This component (referred to hereafter as ‘general cognitive performance’) accounted for 64.6% of the total variance in task performance (Extended Data Table 1), a substantially higher proportion than previous cognitive task batteries on other species^{18–22}. Group size was also the strongest predictor of PC1 (Table S6, Fig. 2). To confirm whether our tasks provide robust measures of individual cognitive performance, we ran a second battery of cognitive tasks two weeks later using causally identical, but visually distinct tasks (see methods). Individual performance was highly repeatable in all four tasks: inhibitory control ($r = 0.806$, $P < 0.0001$), associative learning ($r = 0.97$, $P < 0.0001$), reversal learning ($r = 0.975$, $P < 0.0001$) and spatial memory ($r = 0.932$, $P < 0.0001$) (Extended Data Table 2).

To examine the development of the group size-cognition relationship, we conducted repeated testing of juveniles at 100, 200 and 300 days post-fledging. There was no evidence of general cognitive performance at 100 days post-fledging (see discussion in Supplementary Information), however, much like adults, there was strong evidence for general cognitive

performance at 200 days (PC1 accounted for over 64% of total variance in task performance, Extended Data Table 3, Table S7) and 300 days post-fledging (> 80% of total variance explained by PC1, Extended Data Table 4, Table S8). There was no relationship between group size and cognitive performance at 100 days (Tables S9-10), but PC1 was strongly positively correlated with group size at 200 and 300 days (Fig. 3, Tables S11-12; see Supplementary Information for discussion of influential data points). When analysed longitudinally, an interaction between age tested and group size was the best predictor of cognitive performance (Extended Data Fig. 2, Tables S13-S18).

The emergence of a positive association between group size and cognitive performance through early life supports the possibility that living in large groups helps to drive cognitive development. Manipulations of group size would be required to demonstrate an unequivocal causal effect, which in wild populations may raise virtually insurmountable logistical and ethical challenges (see discussion in Supplementary Information). Our analyses do, however, allow us to address key alternative explanations. First, the elevated cognitive performance of birds in large groups is unlikely to be explained by reduced nutritional constraints on cognitive development²³ because we found no effect of group size on offspring provisioning rates (Table S19), and no relationship between body size and cognitive performance in either adults or juveniles (Tables S1-S4, and S9-S12). We also found no relationship between foraging efficiency and cognitive performance in adults (Tables S1-S4; foraging efficiency data were not available for juveniles). Second, positive effects of group size cannot result from a reduced need for vigilance or reduced neophobia: we recorded no antipredator behaviour during any task presentations and neophobia was unrelated to performance on any tasks, except juveniles' performance on the spatial memory task at 100 days post-fledging (adults: Tables S1-S4; juveniles: Tables S9-S12). There was also no relationship between group size and the time test subjects spent interacting with tasks (see discussion in Supplementary Information). Third, a link between cognitive performance and group size could potentially arise if magpies preferentially joined groups containing individuals with similar traits, but over four years of life-history data collection provide no evidence of such social assortment (see discussion in Supplementary Information). Moreover, we found a clear difference in the frequency distribution of cognitive phenotypes between small and large groups (Extended Data Fig. 3), so it is not simply the case that larger groups have a wider distribution of cognitive phenotypes, and are therefore more likely to contain some high performing individuals by chance. Instead, we propose that, as suggested by captive studies^{13,15}, living in larger groups presents wild animals with information-processing challenges that promote the development of cognitive traits. Determining precisely what those challenges are is a clear priority for future research. An important next step will be to determine whether individual cognitive development is specifically linked to the quantity and quality of their relationships within their social networks, as might be expected if the need to establish and maintain multiple relationships within groups places cognitive demands on individuals³.

To determine whether the group-size dependent cognitive variation we have identified may be subject to selection, we examined the relationship between individual cognitive performance and three measures of reproductive success. General intelligence has been linked to fitness-related traits in humans²⁴, but few studies have examined the fitness

consequences of cognitive variation in wild animals¹¹, and the two that used rigorous psychological test batteries found no effects^{19,25}. In our magpie population, exceptionally high rates of extra-group paternity²⁶ mean that we are only able to reliably identify the mother of the brood (female reproductive skew in our population is low, and all females attempt to breed). Variation in female reproductive success was strongly linked to cognitive performance: general cognitive performance and foraging efficiency were the best predictors of the average number of hatched clutches per female per year (Fig. 4a-b, Table S20), and general cognitive performance alone was the best predictor of the average number of fledglings produced and the average number of fledglings surviving to independence per female per year (Fig. 4c-d, Tables S21-22). These effects were independent of group size (Tables S20-S22) indicating that fitness benefits arise as a direct consequence of elevated cognitive performance and are not simply the result of non-cognitive advantages of living in larger groups. These results provide the first evidence for a potential selective benefit of high levels of general cognitive performance in a wild population of nonhuman animals. Precisely how these benefits arise, and whether elevated cognitive performance incurs any costs²⁷, has yet to be determined. General cognitive performance and foraging efficiency are not correlated in female magpies ($r = 0.06$, $P = 0.791$, $n = 22$), but it is possible that cognitively adept females may boost their reproductive success through improvements not in the quantity, but in the quality or variety of food given to offspring²⁸. Additional, non-mutually exclusive explanations for the relationship between cognition and reproductive success could include enhanced abilities to defend young by avoiding inter- and intra-specific conflict²⁹, or heritable cognitive abilities that promote offspring survival³⁰. It is also possible that the fitness benefits of cognitive performance may account for the group size-cognition relationship, if females with elevated cognitive performance produce large numbers of cognitively adept offspring. However, this explanation is unlikely given that group size is stable over time (see methods), and the extraordinarily high rates of extra-group paternity²⁶ are likely to preclude substantial genetic differentiation between groups.

Since its inception, the Social Intelligence Hypothesis has focused on cognitive differences between species resulting from selection in response to the challenges of social life. Our results indicate that social factors can also have developmental effects on cognition *within* species, with important consequences for individual fitness. In summary, we have shown that wild Australian magpies living in larger groups show elevated cognitive performance, which is in turn associated with increased reproductive success. The association between group size and cognition emerges through early life and cannot be explained by food intake, body size, neophobia, attention to tasks or social assortment. While we cannot rule out the possibility that some other, unmeasured factor could play a role in driving the relationship, our findings strongly suggest that the social environment has developmental effects on fundamental, domain-general cognitive traits. Furthermore, we provide rare evidence that cognitive performance provides benefits for female reproductive success. Recent comparative studies have brought into question the notion that variation in social structure drives cognitive evolution^{9,10}. However, our work highlights the critical importance of considering intra-specific variation, which is typically overlooked by comparative analyses. Together our results point to a major role for the social environment in driving both the development and evolution of cognitive traits.

Methods

Study site and species

The study took place in Guildford, Western Australia, between September 2013 and February 2016. The study population consists of 14 groups of ringed, habituated Australian magpies (Western Australian subspecies *Cracticus tibicen dorsalis*), ranging in size from 3-12 individuals (for composition of study population see Table S23). The Western Australian subspecies breeds cooperatively, and lives in territorial groups where adult group size remains stable (individuals within our study population have remained in the same group since research commenced in 2013, and there have been no recordings of ringed birds moving between groups)^{26,31}. Individuals exhibit a range of cooperative behaviours such as territory defence and alloparental care³². Reproductive skew among females is very low, with all adult females typically attempting to breed each year³³, but extra-group paternity is the highest recorded for any bird species (>82%)²⁶, indicating high gene flow between groups. All of the study population's group territories are located in urban parklands. Although individuals have access to food from anthropogenic sources, it is worth noting that all territories cover similar habitats and none contain dumps or landfills that could provide a glut of food sources.

The majority of birds within our study population are colour-ringed and habituated to close human observation, allowing us to present cognitive tests to most individuals. Individuals are trained to hop onto electronic top-pan scales in return for a crumb of mozzarella cheese, allowing us to collect daily records of individual body mass. Mozzarella was also used as the food reward in the cognitive test battery. Weekly behavioural focal follows are carried out on all individuals in the study population³³, from which foraging efficiency is calculated (defined as the mass of food [in grams], caught per foraging minute; biomass of food items was calculated by Edwards³³).

Adult cognitive test battery

We carried out a battery of cognitive tests on 56 adult Australian magpies. The battery consisted of four tasks designed to measure inhibitory control, associative learning, reversal learning, and spatial memory (Extended Data Fig. 1a-c). All individuals were tested on the tasks in this order. We chose these tasks because (i) they target well-understood and widely studied cognitive traits spanning cognitive domains^{11,20,34} and (ii) they are likely to be highly ecologically relevant: spatial memory is likely to be important in remembering locations of resources and territory boundaries³⁵, while associative and reversal learning enable the acquisition and flexible readjustment of predictive contingencies between cues in the environment, including learning from conspecifics' behaviour^{11,34,36,37}. Finally, inhibitory control, the ability to inhibit prepotent responses, has been implicated in adaptive decision-making in both social and asocial contexts^{16,25,38}.

Inhibitory control—To quantify individuals' ability to inhibit ineffective prepotent responses towards food, we presented individuals with a detour reaching task²⁵. This consisted of a transparent open-ended cylinder (13cm length, 5cm diameter, Extended Data Fig. 1a, Extended Data Fig. 4a) in which a food reward was placed in the centre. Test

subjects were presented with the task such that the open ends of the cylinder were facing away from the individual's direction of gaze. A trial was deemed successful if the test subject inhibited the prepotent response of pecking the transparent cylinder, and detoured around to the open ends of the cylinder to gain access to the food reward. Once an individual successfully detoured to the open ends of the cylinder without pecking the transparent walls three times in a row, it was considered to have passed the inhibitory control task. The number of trials taken to pass was the measure of success. Trials were carried out at one-minute intervals with a maximum of 10 trials, and when possible, all trials were carried out on the same day. Individuals that failed to pass were assigned the maximum score of 10 for statistical analyses. Other studies using the detour-reaching task commonly include a training phase in which test subjects are presented with an opaque tube before being exposed to the transparent tube (e.g. 25). We did not include the opaque phase in our study because it generates difficulties in interpretation: success in the transparent condition could be linked to inhibitory control, or could result from the continued application of a learned rule: pecking at the open ends of the cylinder was rewarded in the opaque condition, so individuals may persist with this behaviour in the transparent condition.

Associative learning—To test associative learning, we used a colour discrimination task consisting of a wooden foraging grid (31 x 9 x 4cm) containing two wells (3.5cm diameter, 2.5cm deep, Extended Data Fig. 1b). The presence of only two wells allowed experimental trials to be carried out quickly, reducing the chance of non-focal birds detecting and approaching the task. The wells were covered with PVC lids that fitted exactly into the wells, and were held in place by elastic bands that were threaded through drilled holes in the lids and fastened to either side of the well (Extended Data Fig. 5). This created an axis on which the lids would swivel when pecked. Birds were first trained to search the wells using a shaping procedure similar to that of Boogert *et al.* 39: magpies could gain access to a food reward (a small amount of grated mozzarella cheese) by first being exposed to the wells without any lids covering them, secondly with the lids partially covering the well, and thirdly with the lid fully covering the well. Lid colour in the training phase was yellow, a colour not used in any of the experimental trials. Once a bird had successfully searched the wells when fully covered by lids three times in a row, it moved onto the experimental trials of the associative learning task.

During experimental trials the wells were covered by either a dark blue or light blue lid. One of these two colours was randomly assigned to be the rewarded colour for each of the test subjects. We used dark and light shades of one colour, rather than distinct colours (e.g. red vs yellow), in order to minimise any potential effects of past experience with particular colours on task performance⁴⁰. Following Shaw *et al.* 20, test subjects were allowed to search both wells in the first trial to demonstrate that only one of the wells contained a food reward. In all subsequent trials, the bird was only allowed to search one well before the task was removed. Test subjects had a maximum of one minute to complete the task. There was a minimum interval of one minute between trials (mean + SD = 1.06 ± 0.35 minutes; range = 1-6 minutes), and a maximum of 50 trials per individual per day; differences in inter-trial interval were unrelated to group size (Spearman's correlation, $r_s = 0.048$, $P = 0.121$, $n = 1027$ trials). If the maximum number of trials was reached on one day, trials were continued

the following day. To ensure that colour was the cue being associated with a food reward, rather than location, the position of the baited well was pseudo-randomised and was never on the same side of the foraging grid for more than three consecutive trials. Furthermore, both wells were wiped with cheese to control for olfactory cues. Following Shaw *et al.* 20, an individual was considered to have passed the task when it pecked the rewarded colour in at least 10 out of 12 consecutive trials (10/12 correct represents a significant deviation from random binomial probability; binomial test: $P = 0.039$). The number of trials taken to reach this criterion (including the final 12 trials) was the associative learning score.

Reversal learning—24hrs after the completion of the associative learning task, individuals were tested on a reversal learning task. The same foraging grid was presented; the only difference being the colour of the rewarded lid was reversed from that of the associative learning task. Otherwise the experimental protocol and the criteria for passing were exactly the same as the associative learning task described above.

Spatial memory—The spatial memory task consisted of a wooden foraging grid (40 x 36 x 4.5cm) containing 8 wells (3.5cm diameter, 2.5cm deep). The wells were equidistant from one another (6 cm between wells) and were arranged in three rows, with the first containing 2 wells, the second 4, and the third 2 (Extended Data Fig. 1c). The wells were covered with light blue lids exactly like those used in the associative and reversal-learning experiments, so no training phase was necessary. Following Shaw *et al.*'s 20 protocol, the experiment consisted of five phases. One of the 8 wells was randomly chosen to be the baited well containing a food reward in all phases of the experiment. The first phase was a “baseline” trial whereby individuals searched the foraging grid for the baited well. Once the test subject had located and eaten the food reward, the foraging grid was removed. Five minutes after the baseline trial, the second “training” phase was carried out, whereby the same well was baited, and the test subjects had to search for the food reward again. The third and fourth phases were test trials where subjects were presented with the foraging grid 24hrs and 48hrs after the training phase respectively. The cumulative number of wells searched before locating the rewarded well in the third and fourth phases of the experiment was the spatial memory score, thus larger scores indicate poorer performance. To control for olfactory cues the foraging grid was presented a fifth time as an unrewarded “probe trial”²⁰ (five minutes after the 48hr post-training phase trial), in which the grid was rotated 180 degrees, and without a baited well. The foraging grid would appear identical to the magpie, but the position of the previously baited well would be on the opposite side of the grid compared to the other phases of the experiment. If the test subject had remembered the location of the rewarded well in the experimental phases, one would predict that it would search the well opposite the previously baited well. If the test subjects were using olfactory cues to locate the rewarded well one would predict that the previously baited well would be searched first. To investigate if birds were using olfactory cues to locate the food reward a paired t-test was carried out to see if there was a difference between the number of wells searched in the 48hr post-training phase trial and the fifth trial (see supplementary discussion for results). The number of wells searched in the fifth presentation did not count towards the spatial memory score.

To ensure that we tested individual performance, and to control for the potentially confounding effect of social learning or social interference, all trials were carried out in conditions as close as possible to social isolation. This was achieved by ensuring that no other birds were within 10m of the bird being tested. This was possible as magpies often forage over 10m away from each other. If another bird approached during an experimental trial, the trial was discontinued. To investigate whether individual performance was affected by social learning, we included “test order” as an explanatory term in the analyses investigating factors affecting performance. This allowed us to verify that individuals tested later within a group (who could therefore have had opportunities to observe previous group members being tested) did not perform better than those tested earlier. Tasks were placed directly in front of the test subjects. Experiments were run between 0500 and 1000 hours and were recorded live by the observers (B.J.A. and E.K.E.). One observer recorded individual performance, whilst the other recorded neophobia (defined as the time elapsed between the test subject first coming within 5m of the apparatus and first touching the apparatus), the time spent interacting with the task, and antipredator behaviour within the group.

Individual consistency in adult performance - repeatability testing

Apparent individual differences in cognitive performance in a single round of testing could simply result from stochastic variation or extraneous confounding variables¹¹. To determine whether individuals were consistent in their performance, we carried out a second test battery two weeks after the first test battery to test the repeatability of adult cognitive performance. To ensure that individuals could not simply use memory of visual cues from the first round of testing to solve tasks in the second round, we changed the visual appearance of each task, while keeping the causal structure of the task the same. In the associative and reversal learning task the colour of lids was changed to dark green and light green. In the spatial memory task the location of the rewarded well was changed from the first test battery. In the inhibitory control task, rather than using an open ended cylinder, we presented food rewards behind a transparent curved wall (30cm length, 10cm height, Extended Data Fig. 4d). Other than these changes in the appearance of the tasks, the protocol and criteria for passing were exactly the same as the first cognitive test battery.

Juvenile cognitive test battery

Juveniles were presented with a battery of four cognitive tasks at three ages: 100, 200 and 300 days post-fledging (Extended Data Fig. 4). Cognitive testing commenced at 100 days post-fledging because by this stage individuals spend the majority of their time foraging independently³¹. The same four cognitive traits (inhibitory control, associative learning, reversal learning, and spatial memory) were quantified at each age by presenting individuals with cognitive test batteries containing causally identical but visually distinct versions of each of the four tasks (Extended Data fig. 4). This ensured the same cognitive traits were tested at each age, whilst making sure the tasks were not the same in appearance, minimising the potentially confounding effect of memory.

To quantify inhibitory control at 100 days post-fledging we presented individuals with the same detour reaching task used in the adult cognitive test battery (Extended Data Fig. 1a,

Extended Data Fig. 4a), and used the same experimental protocols and the same criteria for passing the task. At 200 days post-fledging, rather than using a transparent open ended cylinder, food rewards were presented behind a transparent curved wall (32cm length, 12cm height, Extended Data Fig. 4d). At 300 days post-fledging individuals were presented with a detour reaching task consisting of a transparent “umbrella”, whereby a food reward could be accessed by detouring underneath the transparent Perspex dome (55cm circumference, 8cm height, Extended Data Fig. 4g). Regardless of the differences in appearance, the criterion for passing the inhibitory control tasks at 200 and 300 days post-fledging was exactly the same as that of the first detour-reaching task presented at 100 days post-fledging.

Spatial memory was quantified at 100 days post-fledging by presenting individuals with a wooden foraging grid (40 x 26 x 4cm) containing six wells (3.5cm diameter, 2.5cm deep) covered with lids exactly like those used in the associative and reversal learning tasks (Extended Data Fig. 4c). One of the six wells was randomly assigned to be the rewarded location for all phases of the experiment. The spatial memory experiment consisted of two phases; first the grid was presented in a “baseline trial” whereby individuals were able to search for the rewarded location. A “memory trial” was carried out five minutes later, where individuals were presented with the foraging grid in the exact same arrangement a second time. The number of wells searched before locating the food reward in the memory trial was the spatial memory score. At 200 and 300 days post-fledging the same experiment was carried out, although we ensured that a different well was randomly assigned as the rewarded location (Extended Data Fig. 4f and 4i).

We were unable to quantify individual performance in associative and reversal learning at 100 days post-fledging because individuals took a prohibitive amount of trials to complete the tasks (no individuals passed within 20 trials). Associative learning was quantified at 200 days post-fledging by presenting test subjects with a wooden foraging grid (41 x 35 x 4cm) containing 20 wells (3.5cm diameter, 2.5cm deep), covered with 20 plastic lids; 10 a light blue colour and 10 a dark blue colour (Extended Data Fig. 4e). Wells covered with lids of one colour were randomly assigned to be rewarded for the duration of the trials, whereby when pecked, a food reward could be accessed. Test subjects were considered to have passed the associative learning task when they chose the rewarded well in eight out of the first nine wells searched; this represents a significant deviation from binomial probability (binomial test: $P = 0.039$). The number of trials taken to reach this criterion was the associative learning score. Individuals received a maximum of 10 trials; those that failed to pass were assigned a score of 10 for statistical analyses. At 300 days post-fledging rather than light and dark blue lids, light and dark green lids were used (Extended Data Fig. 4h). Other than the change in lid colour, all protocols and criteria for passing remained the same.

At each developmental stage (200 and 300 days post-fledging), reversal learning was quantified 24 hours after the successful completion of the associative learning task. The protocol and criteria for passing the reversal learning tasks were exactly the same as the associative learning task, except that the previously unrewarded colours were now rewarded.

As with the cognitive testing carried out on adults, all trials were carried out on juveniles in isolation. This was achievable because by 100 days post-fledging, juveniles are foraging independently.

Life-history data collection

To obtain measures of reproductive success for individual birds, we collected life-history data on the study population over three years. This was collected through a combination of behavioural focal follows on individuals, brood observations, and adlib data collected while watching the whole group (for details see 31 and 33). The extensive life history database developed from these observations allowed us to determine the number of hatched clutches, the number of nestlings that fledged, and the number of fledglings surviving to independence for each adult female in the study population per annum. In addition, the behavioural focal observations, brood observations, and adlib data allowed us to quantify the amount of food adults provisioned to young. Fledglings were considered to have survived to independence when they reached three months post-fledging. At this age, magpies forage independently and are fed by adults infrequently³¹. In addition to these three proxies of fitness, we also recorded the number of breeding attempts by females - a breeding attempt was considered to have occurred if a female was observed incubating on a nest. The mother was assumed to be the bird incubating at the nest (there is no evidence of egg-dumping or shared incubation in this subspecies, so there was only ever one female incubating at a given nest). Groups were visited at least once a week during the breeding season, providing accurate measures of the number of breeding attempts made per female, and accurate hatch and fledge dates for all nests. Clutches were considered to have hatched when adults started bringing food to the nest, or if we could see young in the nest. As many nests were upwards of 20m high, we were unable to accurately determine clutch size to use as an additional measure of reproductive success.

All methods were performed in accordance with the University of Western Australia's guidelines and regulations, and were approved by the University of Western Australia Animal Ethics Office (ref: RA/3/100/1272).

Statistical analyses

Adult cognitive performance—To determine the factors influencing individual variation in cognitive ability we analysed cognitive performance using generalised linear mixed models (glmm) with either a poisson distribution with a logarithmic link (inhibitory control), or a negative binomial distribution with a logarithmic link to account for over-dispersion (associative learning, reversal learning and spatial memory). Cognitive performance was measured as the number of trials taken to pass the task. In addition to the potential cognitive demands of living in larger social groups, it is possible that indirect effects of group size on energy intake and task attention could generate group size effects on cognitive performance^{41,42}. We therefore included neophobia (defined as the time taken to interact with the task once being within 5m of it), body mass, and foraging efficiency as explanatory terms in the analysis, as well as sex, the sex ratio of males to females in the group, the order tested within the group, and group size. Group identity was included as a random term in all models.

To determine whether body condition (body mass, accounting for skeletal size) could explain variation in cognitive performance, we included mass (in grams) and tarsus size (in mm; a common measure of skeletal size in birds) as covariates in an additional analysis on a subset of individuals for which both of these morphometric measures were available ($n = 27$). Dominance status was not included as an explanatory variable as there is not a clear dominance hierarchy within magpie groups. Adult age and immigration status were not included as explanatory variables because the fledge date and natal origins of some of the adults in our population is unknown (Australian magpies are very long-lived, living up to 25 years in the wild³²). We note that among the birds whose complete life-history is known ($n = 19$ individuals), there has been no movement between groups.

We analysed our data using a model selection process; terms were ranked in order of their corrected quasi-information criterion (QICc) values (the lowest QICc value has the greatest explanatory power⁴³). If a term was more than two QICc units smaller than any other term, then this was judged to explain the observed relationship in the data better than any other term. If there was more than one term with $QICc < 2$ from the ‘best’ term, had confidence intervals that did not intersect zero, and explained more variation than the basic model (the model containing no predictors, just the constant and the random terms), then model averaging was carried out on this “top set” of models *sensu* Symonds and Moussalli⁴⁴. All statistical analyses were carried out using IBM SPSS Statistics software (version 22).

To examine the relationship in performance across tasks, we conducted Spearman’s rank pairwise correlations between all four tasks. To determine if a general cognitive factor explained cognitive performance across all four tasks, we performed a Principal Components Analysis (PCA) with a varimax rotation. Only principal components with an eigenvalue > 1 were extracted from the analysis. A general intelligence factor has been argued to exist when all tasks load positively onto the first principal component and explain $> 30\%$ of total task variance²². Following Shaw et al. 20, to assess whether the tasks loaded onto the first principal component by chance we compared the mean and standard deviation of the first component factor loadings to the 95% confidence intervals of the means and standard deviations of the first component factor loadings from 10,000 simulations. For each simulation, performance within each task was randomised between individuals (using the `randomizeMatrix` function in the `picante` R package⁴⁵), a PCA was performed, and the mean and standard deviation of the first component factor loadings were obtained. The 95% confidence intervals were then calculated from the stored means and standard deviations from all the simulations.

Statistical analyses used to calculate estimates of repeatability in cognitive performance between the first and second cognitive test batteries were carried out in R (version 3.1.1, <http://www.r-project.org>) with the `rptR` package⁴⁶ using a linear mixed model repeatability estimate, with a restricted maximum likelihood function (`reml`).

Juvenile cognitive performance—A series of glms were carried out to determine factors affecting cognitive performance in each of the tasks. Model selection (using the same approach as for analyses on adult cognitive performance) was then used to determine the most significant predictors of performance in each of the cognitive tasks⁴³.

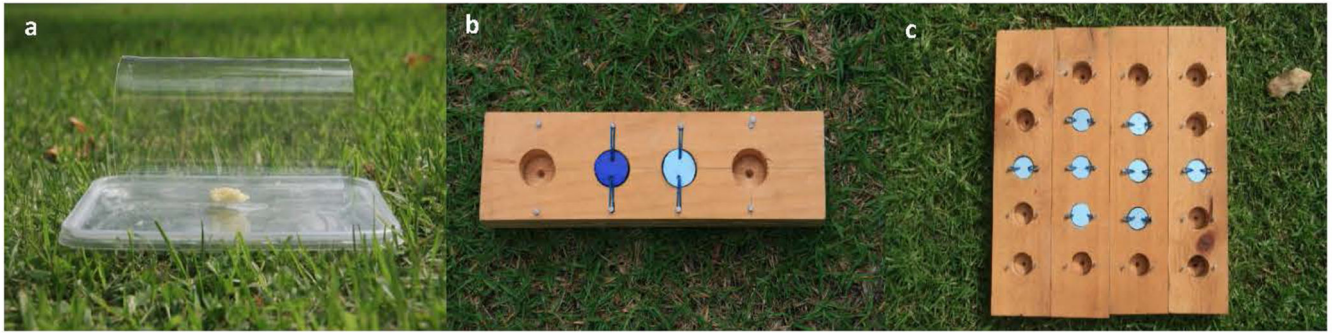
At 100 days post-fledging, the response terms used were cognitive performance; in the detour reaching task this was the number of trials until passed, and in the spatial memory task it was the number of wells searched. As these were count data, generalised linear mixed models with a poisson distribution were used. The relationship between performance in the detour-reaching task and the spatial memory task were examined using a spearman rank correlation. At 200 and 300 days post-fledging, we found evidence of general cognitive performance in juvenile magpies (Extended Data Tables 3 and 4); this parameter was therefore used as the response term for analyses investigating factors affecting cognitive performance at 200 and 300 days post-fledging.

Explanatory terms included in the models were neophobia, body mass, what stage of the breeding season (early or late), the presence or absence of siblings (from the same brood), group size, and the sex ratio of adult males to females in the group. We were unable to include provisioning rate from adults to fledglings as an explanatory term in analyses as this data was only available for a small subset of individuals. Group ID was included as a random term in all models.

Factors affecting performance across all ages were analysed for each of the four cognitive traits quantified, using generalised linear mixed models. Four separate analyses were carried out, with cognitive performance used as the response term. An additional two analyses were carried out, firstly to determine factors affecting performance across all ages for both inhibitory control and spatial memory (associative and reversal learning were omitted from this analysis as we only quantified performance at 200 and 300 days post-fledging in these traits). Secondly, we investigated factors affecting general cognitive performance measured at 200 and 300 days post-fledging. Group ID and individual ID were included as random terms. Explanatory terms included were those used for the previous analyses. A model selection approach was used to determine the most significant terms affecting performance.

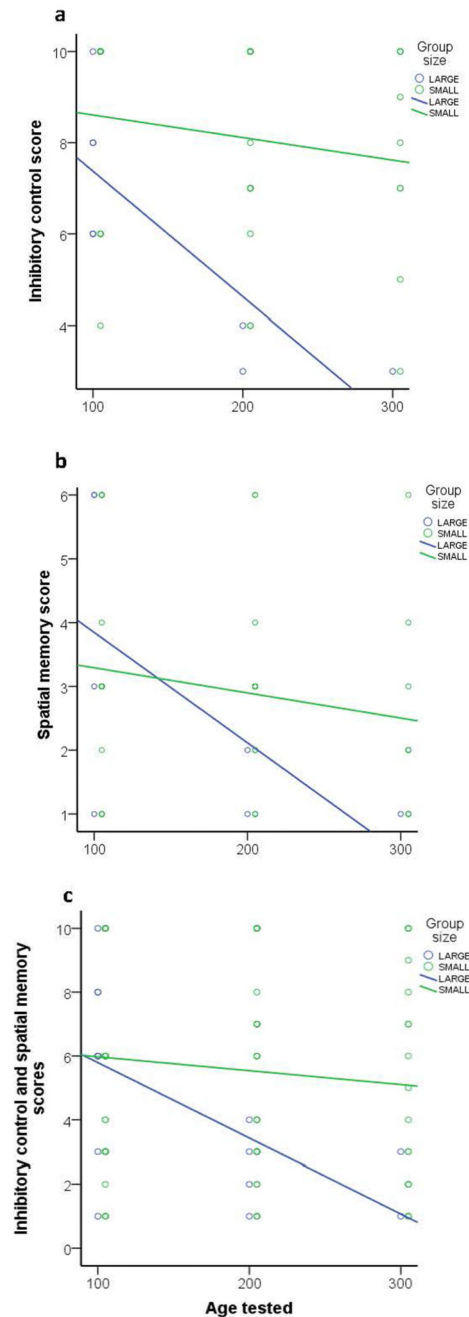
Relationship between cognitive performance and measures of reproductive success—We carried out three separate analyses to determine the factors affecting three measures of reproductive success: the average number of hatched clutches per year, the average number of nestlings fledged per year, and the average number of fledglings surviving to independence per year. We carried out glmms, with the reproductive success measure as the response term, and group ID included as a random term. Explanatory terms included in the analyses were body mass, foraging efficiency, group size, the sex ratio of the group, and general cognitive performance. General cognitive performance was used as an explanatory term for cognitive performance because the PCA revealed robust evidence for its existence within females (PC1 accounted for >70% of total variance in female task performance, Table S25). We did not include age because we do not know the exact fledge date for the majority of adult females in the population.

Extended Data



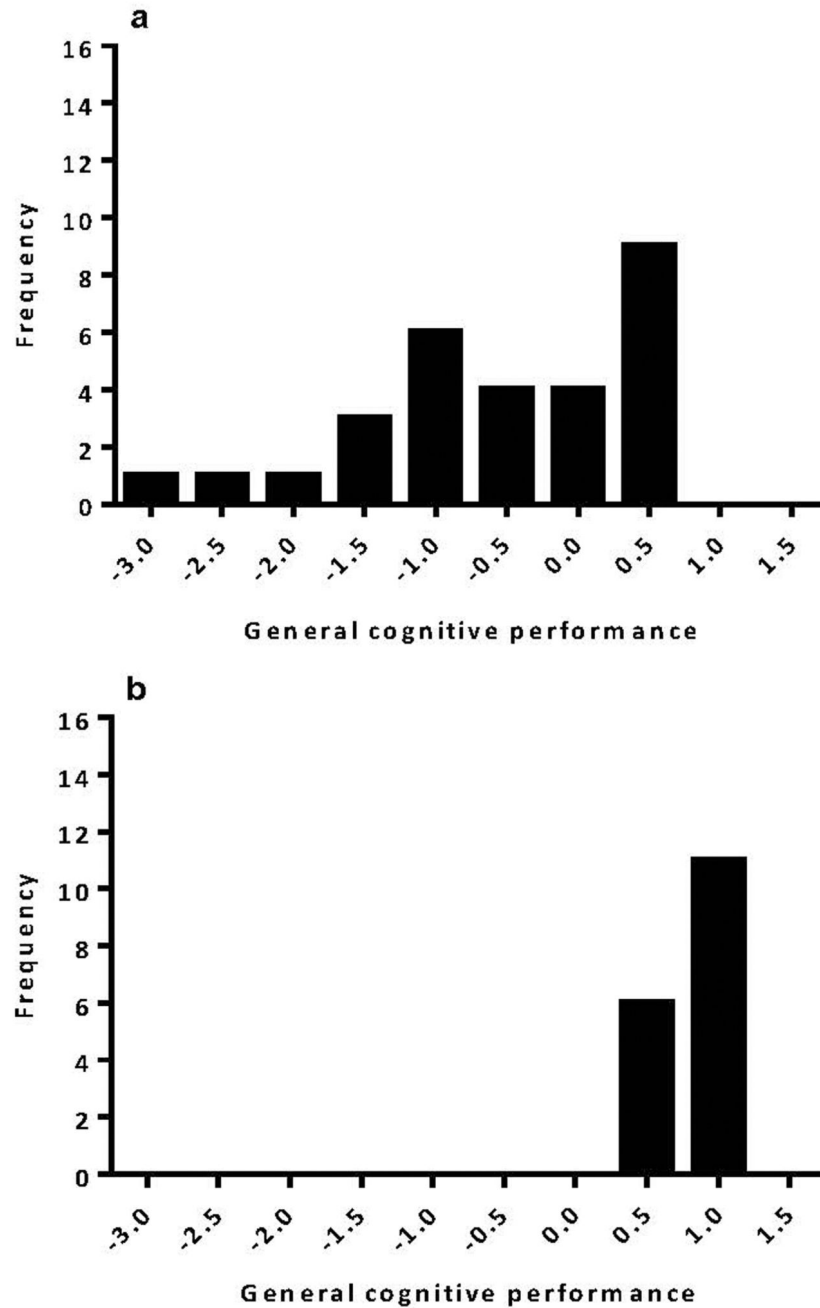
Extended Data Fig. 1. Adult cognitive test battery.

The cognitive test battery used to quantify individual variation in (a) inhibitory control (b) associative and reversal learning (c) spatial memory.



Extended Data Fig. 2. Developmental trajectory of cognitive performance.

The developmental trajectory of Australian magpies at 100, 200, and 300 days post-fledging in two cognitive traits: (a) behavioural inhibition (n=48 trials) (b) spatial memory (n=46 trials), and (c) behavioural inhibition and spatial memory combined (n=94 trials). Green dots = individuals from small groups (containing 1-7 individuals); blue dots = individuals from large groups (8 individuals). Scores are measured as either the number of trials taken to pass the task, or the number of locations searched, so smaller scores indicate better performance.



Extended Data Fig. 3. Frequency distribution of general cognitive performance in relation to group size.

Frequency distribution of general cognitive performance among individuals in (a) small groups (containing <8 individuals), n=29 individuals, and (b) large groups (>8 individuals), n=17 individuals.



Extended Data Fig. 4. Juvenile cognitive test batteries.

Cognitive test batteries presented to individuals at 100 (**a-c**), 200 (**d-f**), and 300 (**g-i**) days post-fledging, containing four tasks designed to quantify Inhibitory control (**a, d, g**), associative and reversal learning (**b, e, h**), and spatial memory (**c, f, i**). Figure (**b**) is black because individuals were unable to complete the associative and reversal learning tasks at 100 days post-fledging. Red circles indicate that individuals had to search a different location at each age tested in order to obtain the food reward in the spatial memory task.



Extended Data Fig. 5. The lids used in the cognitive tasks.

The lids used in the associative learning, reversal learning, and spatial memory tasks. The lids were held firmly in place by elastic bands, and would swivel when pecked, allowing individuals to search wells for their contents.

**Extended Data Table 1
Principal components analysis (adults).**

Results of the principal components analysis for adult magpies that completed all four tasks. All tasks loaded positively onto the one principal component extracted with an eigenvalue >1 . $n = 46$ individuals.

Task	PC1
Inhibitory control	0.703
Associative learning	0.789
Reversal learning	0.870
Spatial memory	0.841
Eigenvalue	2.582
% of total variance explained	64.56

Extended Data Table 2
Repeatability of cognitive performance.

Estimations of repeatability between the first cognitive test battery and the second cognitive test battery.

Task	Repeatability	SE	Confidence intervals	P
Inhibitory control	0.806	0.049	0.691, 0.882	<0.0001
Associative learning	0.970	0.01	0.946, 0.983	<0.0001
Reversal learning	0.975	0.008	0.954, 0.986	<0.0001
Spatial memory	0.932	0.021	0.879, 0.963	<0.0001

Extended Data Table 3
Principal components analysis (200 days post-fledging).

Results of the principal components analysis for magpies that completed all four tasks at 200 days post-fledging. All tasks loaded positively onto the one principal component extracted with an eigenvalue >1. n = 15 individuals.

Task	PC1
Inhibitory control	0.571
Associative learning	0.916
Reversal learning	0.941
Spatial memory	0.737
Eigenvalue	2.593
% of total variance explained	64.837

Extended Data Table 4
Principal components analysis (300 days post-fledging).

Results of the principal components analysis for magpies that completed all four tasks at 300 days post-fledging. All tasks loaded positively onto the one principal component extracted with an eigenvalue >1. n = 10 individuals.

Task	PC1
Inhibitory control	0.675
Associative learning	0.947
Reversal learning	0.972
Spatial memory	0.957
Eigenvalue	3.215
% of total variance explained	80.363

Supplementary Information

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank Eleanor Russell and the late Ian Rowley for access to their life history records, and for allowing us to continue work on their Guildford magpie population. We thank Amy Wolton for help with fieldwork, Rowan Lymbery for help with statistical analyses, and Neeltje Boogert and Andy Young for helpful comments and discussion. This work was funded by an ARC Discovery grant awarded to A.R.R., A.T. & Matthew B.V. Bell and a University of Western Australia International Postgraduate Research Scholarship and Endeavour Research Fellowship awarded to B.J.A. A.T. received additional support from a BBSRC David Phillips Fellowship (BB/H021817/1).

References

1. Humphrey, N. Growing Points in Ethology. Bateson, PPG., Hinde, RA., editors. Cambridge Univ. Press; 1976. p. 303-317.
2. Byrne, RW., Whiten, A. Machiavellian intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes, and Humans. Clarendon press; 1988.
3. Shultz S, Dunbar RI. The evolution of the social brain: anthropoid primates contrast with other vertebrates. *Proc R Soc B Biol Sci.* 2007; 274:2429–2436.
4. Bond AB, Kamil AC, Balda RP. Social complexity and transitive inference in corvids. *Anim Behav.* 2003; 65:479–487.
5. Street SE, Navarrete AF, Reader SM, Laland KN. Coevolution of cultural intelligence, extended life history, sociality, and brain size in primates. *Proc Natl Acad Sci.* 2017; 114:1–7.
6. Forss SIF, Willems E, Call J, van Schaik CP. Cognitive differences between orang-utan species: a test of the cultural intelligence hypothesis. *Sci Rep.* 2016; 6 30516.
7. Fox KCR, Muthukrishna M, Shultz S. The social and cultural roots of whale and dolphin brains. *Nat Ecol Evol.* 2017; 1:1699–1705. [PubMed: 29038481]
8. Holekamp KE. Questioning the social intelligence hypothesis. *Trends Cogn Sci.* 2007; 11:65–9. [PubMed: 17188553]
9. DeCasien AR, Williams SA, Higham JP. Primate brain size is predicted by diet but not sociality. *Nat Ecol Evol.* 2017; 1:112. [PubMed: 28812699]
10. Sayol F, et al. Environmental variation and the evolution of large brains in birds. *Nat Commun.* 2016; 7 13971.
11. Morand-Ferron J, Cole EF, Quinn JL. Studying the evolutionary ecology of cognition in the wild: a review of practical and conceptual challenges. *Biol Rev.* 2015; 91:367–389. [PubMed: 25631282]
12. Thornton A, Lukas D. Individual variation in cognitive performance: developmental and evolutionary perspectives. *Philos Trans R Soc Lond B Biol Sci.* 2012; 367:2773–83. [PubMed: 22927576]
13. Sallet J, et al. Social Network Size Affects Neural Circuits in Macaques. *Science.* 2011; 334:697–700. [PubMed: 22053054]
14. Kanai R, Bahrami B, Roylance R, Rees G. Online social network size is reflected in human brain structure. *Proc R Soc London B Biol Sci.* 2011; 279:1327–1334.
15. Fischer S, Bessert-Nettelbeck M, Kotschal A, Taborsky B. Rearing-Group Size Determines Social Competence and Brain Structure in a Cooperatively Breeding Cichlid. *Am Nat.* 2015; 186:123–40. [PubMed: 26098344]
16. MacLean EL, et al. The evolution of self-control. *Proc Natl Acad Sci.* 2014; 111:E2140–E2148. [PubMed: 24753565]
17. Plomin R. The genetics of g in human and mouse. *Nat Rev Neurosci.* 2001; 2:136–141. [PubMed: 11252993]
18. Chandra SB, Hosler JS, Smith BH. Heritable variation for latent inhibition and its correlation with reversal learning in honeybees (*Apis mellifera*). *J Comp Psychol.* 2000; 114:86–97. [PubMed: 10739314]
19. Isden J, Panayi C, Dingle C, Madden J. Performance in cognitive and problem-solving tasks in male spotted bowerbirds does not correlate with mating success. *Anim Behav.* 2013; 86:829–838.

20. Shaw RC, Boogert NJ, Clayton NS, Burns KC. Wild psychometrics: evidence for 'general' cognitive performance in wild New Zealand robins, *Petroica longipes*. *Anim Behav*. 2015; 109:101–111.
21. Hopkins WD, Russell JL, Schaeffer J. Chimpanzee intelligence is heritable. *Curr Biol*. 2014; 24:1649–1652. [PubMed: 25017206]
22. Galsworthy MJ, et al. Assessing reliability, heritability and general cognitive ability in a battery of cognitive tasks for laboratory mice. *Behav Genet*. 2005; 35:675–692. [PubMed: 16184494]
23. Buchanan KL, Grindstaff JL, Pravosudov VV. Condition dependence, developmental plasticity, and cognition: Implications for ecology and evolution. *Trends Ecol Evol*. 2013; 28:290–296. [PubMed: 23518414]
24. Plomin R, Deary I. Genetics and intelligence differences: five special findings. *Mol Psychiatry*. 2014; 20:98–108. [PubMed: 25224258]
25. Boogert NJ, Anderson RC, Peters S, Searcy WA, Nowicki S. Song repertoire size in male song sparrows correlates with detour reaching, but not with other cognitive measures. *Anim Behav*. 2011; 81:1209–1216.
26. Hughes JM, et al. High levels of extra-group paternity in a population of Australian magpies *Gymnorhina tibicen*: evidence from microsatellite analysis. *Mol Ecol*. 2003; 12:3441–3450. [PubMed: 14629358]
27. Kotschal A, et al. Artificial selection on relative brain size in the guppy reveals costs and benefits of evolving a larger brain. *Curr Biol*. 2013; 23:168–171. [PubMed: 23290552]
28. Arnold KE, Ramsay SL, Donaldson C, Adam A. Parental prey selection affects risk-taking behaviour and spatial learning in avian offspring. *Proc R Soc B Biol Sci*. 2007; 274:2563–2569.
29. Silk JB, Alberts SC, Altmann J. Social Bonds of Female Baboons Enhance Infant Survival. *Science*. 2003; 302:1231–1235. [PubMed: 14615543]
30. Croston R, Branch CL, Kozlovsky DY, Dukas R, Pravosudov VV. Heritability and the evolution of cognitive traits. *Behav Ecol*. 2015; 26:1447–1459.
31. Pike, KN. How much do helpers help? Variation in helper behaviour in the cooperatively breeding Western Australian magpie. Thesis, University of Western Australia; 2016.
32. Kaplan, G. Australian magpie: biology and behaviour of an unusual songbird. CSIRO Publishing; 2004.
33. Edwards EK, Mitchell NJ, Ridley AR. The impact of high temperatures on foraging behaviour and body condition in the Western Australian Magpie *Cracticus tibicen dorsalis*. *Ostrich*. 2015; 86:137–144.
34. Morand-ferron J. Why learn? The adaptive value of associative learning in wild populations. *Curr Opin Behav Sci*. 2017; 16:73–79.
35. Sherry, DF. Cognitive ecology: the evolutionary ecology of information processing and decision making. University of Chicago Press; 1998. p. 261–296.
36. Lotem A, Halpern JY. Coevolution of learning and data-acquisition mechanisms: a model for cognitive evolution. *Philos Trans R Soc Lond B Biol Sci*. 2012; 367:2686–94. [PubMed: 22927567]
37. Leadbeater E. What evolves in the evolution of social learning? *J Zool*. 2015; 295:4–11.
38. Amici F, Aureli F, Call J. Fission-fusion dynamics, behavioral flexibility, and inhibitory control in primates. *Curr Biol*. 2008; 18:1415–9. [PubMed: 18804375]
39. Boogert NJ, Giraldeau L-A, Lefebvre L. Song complexity correlates with learning ability in zebra finch males. *Anim Behav*. 2008; 76:1735–1741.
40. Rowe C, Healy SD. Measuring variation in cognition. *Behav Ecol*. 2014; 25:1287–1292.
41. Griffin AS, Guez D. Innovative problem solving in nonhuman animals: the effects of group size revisited. *Behav Ecol*. 2015; 26:722–734.
42. Nowicki S, Searcy WA, Peters S. Brain development, song learning and mate choice in birds: A review and experimental test of the 'nutritional stress hypothesis'. *J Comp Physiol A Neuroethol Sensory, Neural, Behav Physiol*. 2002; 188:1003–1014.
43. Burnham, KP., Anderson, DR. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. Springer; 2002.

44. Symonds MRE, Moussalli A. A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike's information criterion. *Behav Ecol Sociobiol.* 2011; 65:13–21.
45. Kembel SW, et al. R tools for integrating phylogenies and ecology. 2014
46. Nakagawa S, Schielzeth H. Repeatability for Gaussian and non-Gaussian data: A practical guide for biologists. *Biol Rev.* 2010; 85:935–956. [PubMed: 20569253]

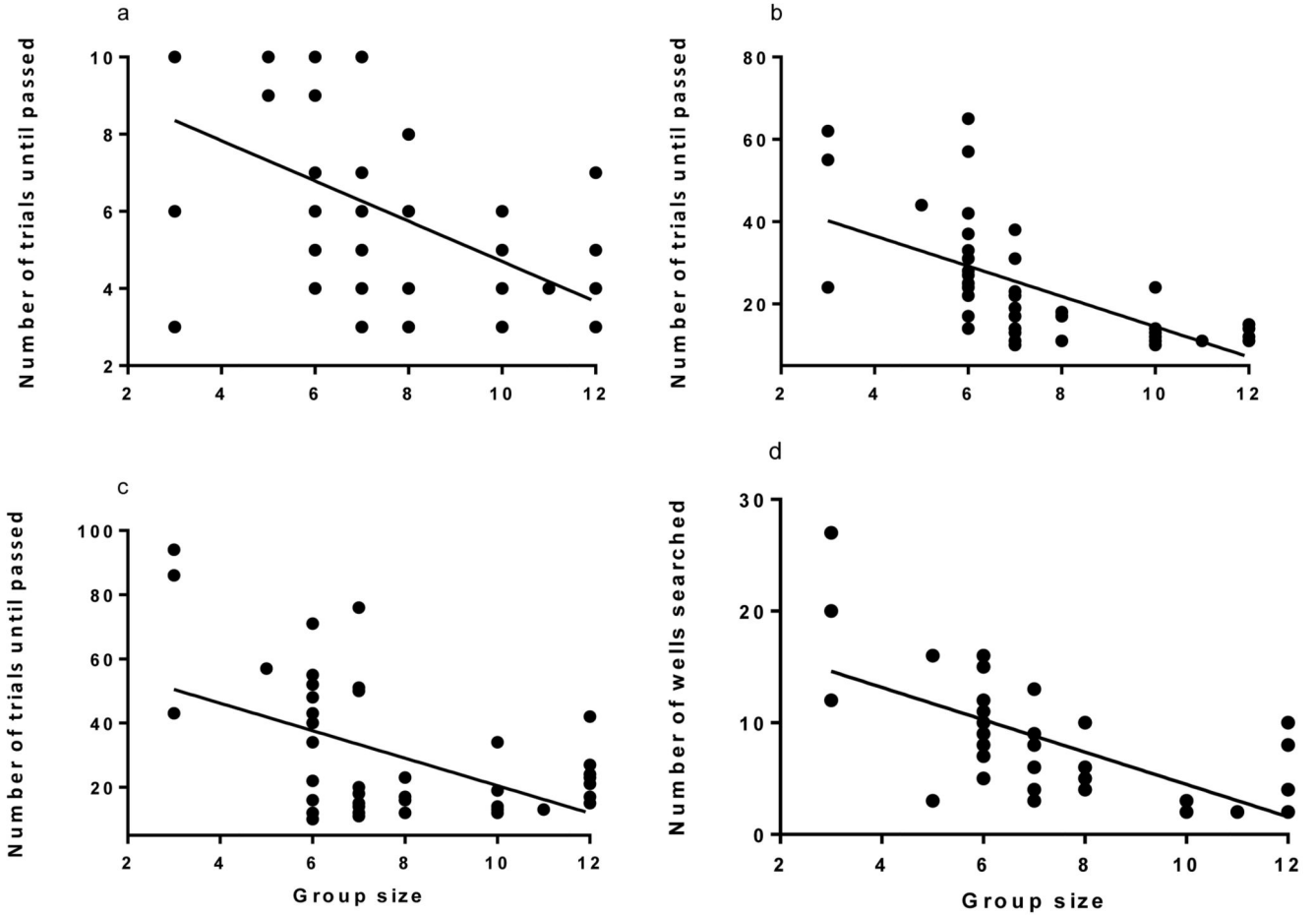


Figure 1. The relationship between group size and cognition. The relationship between group size and cognitive performance in a (a) inhibitory control task, n = 56 individuals, (b) associative learning task, n = 48 individuals, (c) reversal learning task, n = 48 individuals, and (d) spatial memory task, n = 49 individuals. Lines represent best fit. Performance is measured as either the number of trials taken to pass the task, or the number of locations searched, so smaller scores indicate better performance.

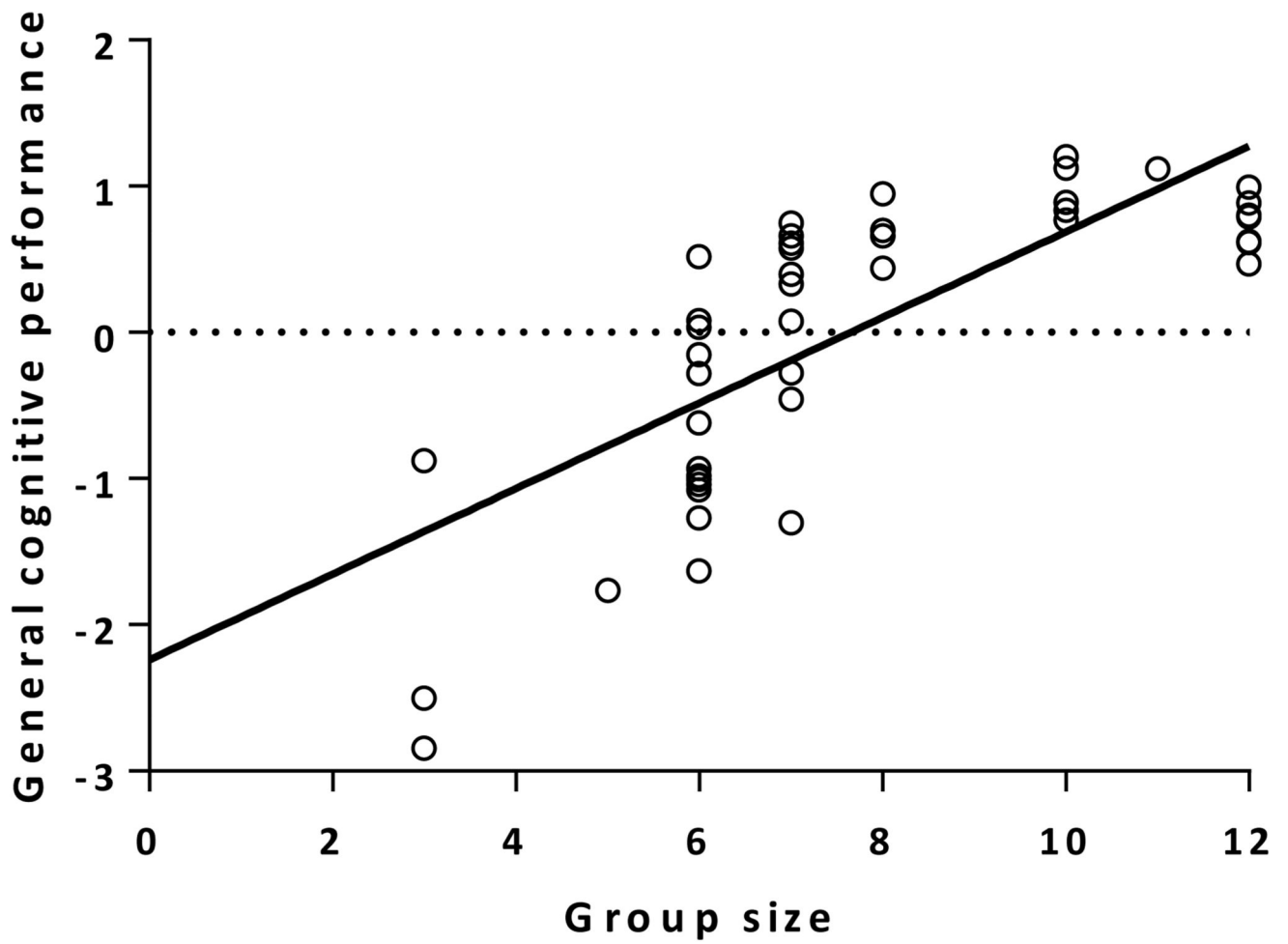


Figure 2. The relationship between group size and general cognitive performance. Individual measures of general cognitive performance derived from principal components analysis. $n = 46$ individuals.

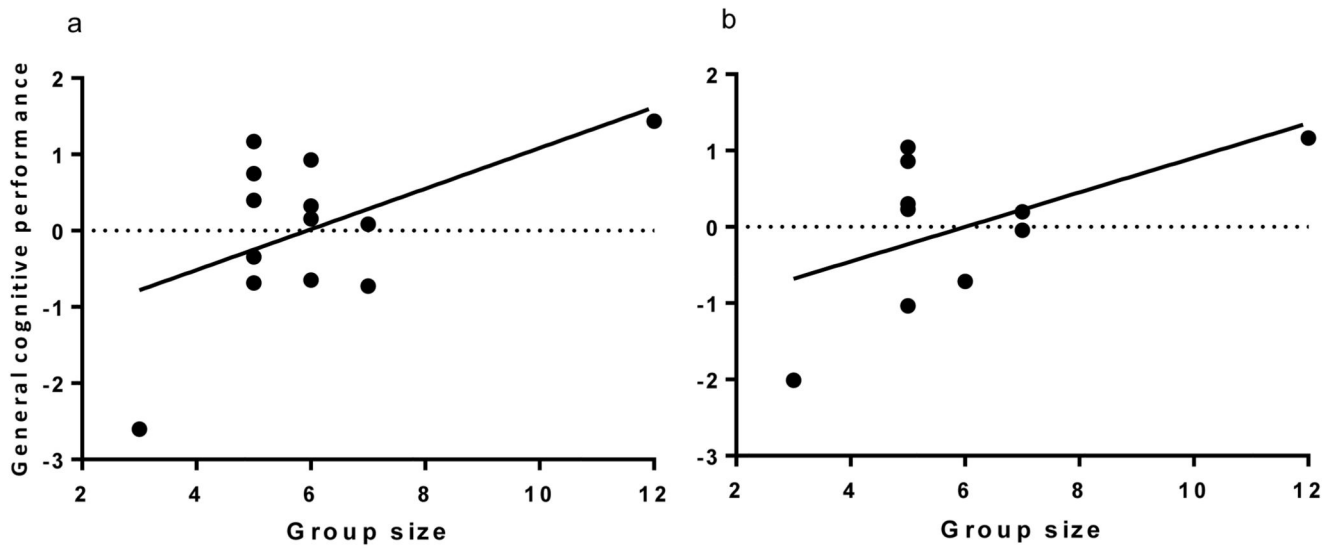


Figure 3. The relationship between general cognitive performance and group size in juveniles. The relationship between general cognitive performance and group size at (a) 200 days post-fledging, n=15 individuals, and (b) 300 days post-fledging, n=10 individuals. General cognitive performance could not be computed at 100 days post-fledging.

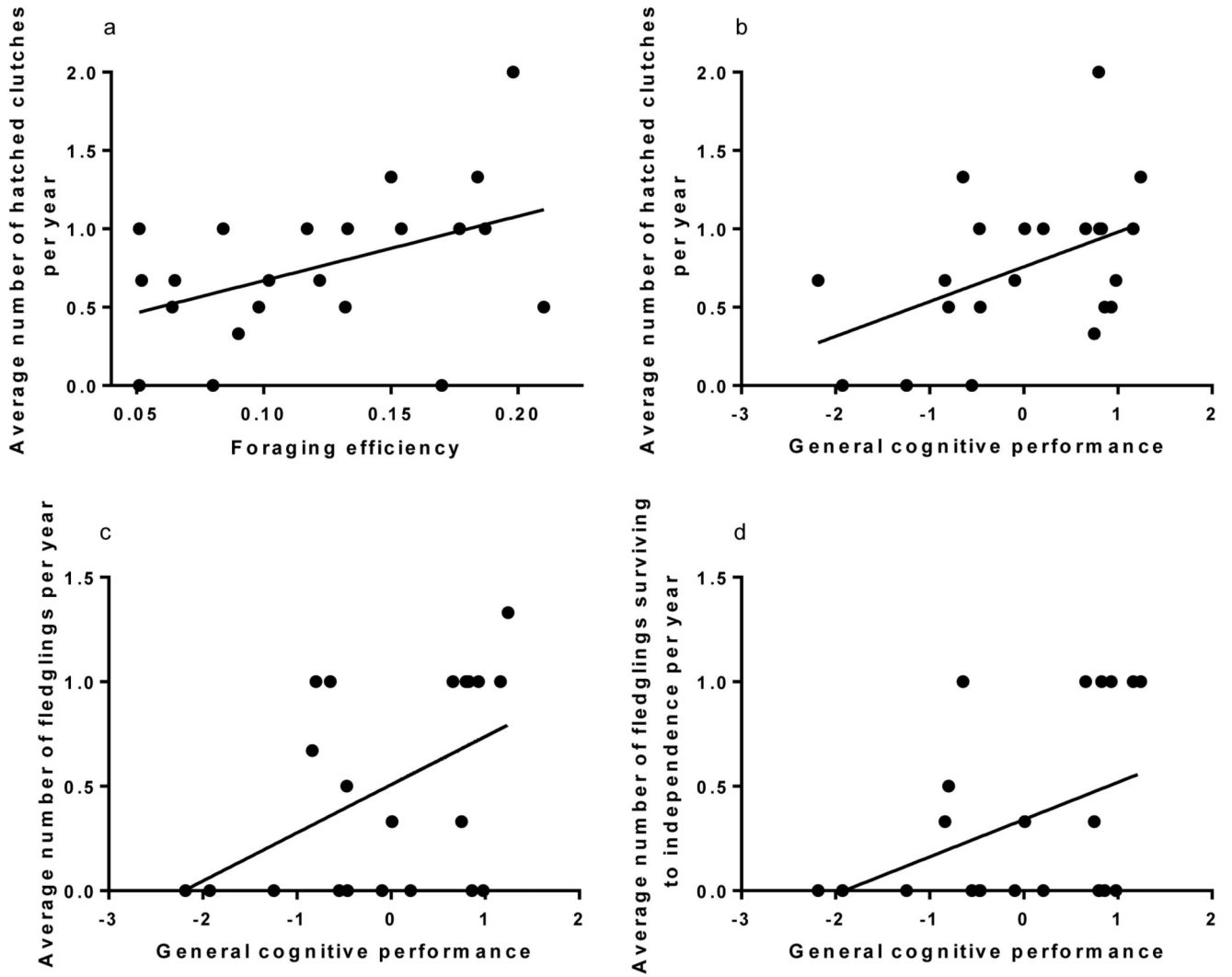


Figure 4. The relationship between female traits and reproductive success.

The relationship between (a) foraging efficiency and the average number of hatched clutches per female per year, (b) general cognitive performance and the average number of hatched clutches per female per year, (c) general cognitive performance and the average number of fledglings per female per year, and (d) general cognitive performance and the average number of fledglings surviving to independence per female per year. $n = 22$ individuals.