Animal Behaviour 196 (2023) 55–81

Contents lists available at ScienceDirect

Animal Behaviour

journal homepage: www.elsevier.com/locate/anbehav

Learning predictably changing spatial patterns across days in a food-caching bird

Lauren M. Benedict [a], *, Virginia K. Heinen [a], Benjamin R. Sonnenberg [a], Eli S. Bridge [b], Vladimir V. Pravosudov [a]

[a] Department of Biology, University of Nevada Reno, NV, U.S.A.
[b] University of Oklahoma, Oklahoma Biological Survey, Norman, OK, U.S.A.

ARTICLE INFO

Article history:
Received 20 December 2021
Initial acceptance 16 February 2022
Final acceptance 26 September 2022

MS. number: A21-00709R

Keywords:
avian cognition
chickadee
cognitive ecology
cognitive flexibility
learning and memory
rule learning
spatial cognition

Environmental variability favours the evolution of learning and memory, influencing not only basic associative learning processes but also more advanced cognitive abilities associated with cognitive flexibility. When environmental conditions change repeatedly and predictably, the ability to learn related patterns and anticipate future changes can be highly adaptive. We tested whether food-caching mountain chickadees, *Poecile gambeli*, from different elevations could (1) successfully learn daily alternating food locations in an eight-position spatial serial reversal task across multiple days and (2) use the daily alternating rule to predict the next day's food location under natural conditions. Chickadees learned the alternating, serial reversal task successfully but birds from high elevations with harsher, less predictable winter environmental conditions performed worse than chickadees from milder, more predictable low elevations. In addition, we found evidence that some birds at low but not high elevation were capable of learning to predict which feeder would provide food the next day after switching food locations just seven times. This behaviour suggests that chickadees recalled which feeder was rewarding the previous day in order to anticipate the feeder that would provide a food reward on the current day. Overall, our study suggests that food-caching chickadees are highly cognitively flexible and show performance consistent with learning the reversal rule and are able to predict a learned switching pattern, but such flexibility may be a trade-off with stronger spatial memories and higher memory load favoured by harsher winter environmental conditions.

In variable environments, learning and memory can be highly adaptive, allowing animals to adjust their behaviour to varying conditions based on previous experiences (Aoki & Feldman, 2014; Dridi & Lehmann, 2016; Dunlap & Stephens, 2016; Stephens, 1991). Learning is most effective when environments are uncertain (i.e. environmental conditions such as precipitation, temperature and food availability change frequently) and previous experience is reliable (i.e. when learned information leads to the best action in future situations; Dunlap & Stephens, 2016; Stephens, 1991). In these environments, using associative learning to remember relationships between stimuli and responses may provide an advantage (Dunlap & Stephens, 2016). Moreover, if environmental variation repeats nonrandomly, then individuals may be able to learn general rules and even learn how to use these rules to predict (or anticipate) future conditions (e.g. Murphy et al., 2008). This rule-learning approach may be more efficient than repeatedly learning and relearning individual associations, as the latter may interfere with learning new information or with retrieving previous memories through proactive or retroactive interference (Shettleworth, 2010; Tello-Ramos et al., 2019).

Regular environmental changes can be associated with multimodal cues, consisting of visual, auditory, spatial or even temporal stimuli. Basic associative learning and memory abilities can be used to make associations with each environmental cue, but flexibly relearning information when conditions change may be difficult due to proactive interference (Anderson & Neely, 1996; Jacoby et al., 2001; Wixted, 2004). With relearned information, multiple memories may become associated with the same cue and subsequently ‘interfere’ with each other during memory recall (Bjork, 1989; Tello-Ramos et al., 2018). The strength of proactive interference appears related to the strength of initial memories (i.e. memory longevity and persistence; reviewed in Tello-Ramos et al., 2019) and how similar the cues and contexts are (Lewis et al., 2013; Rodriguez et al., 1993). As a result, individuals that have better learning and

---

* Corresponding author.
E-mail address: lbenedict@nevada.unr.edu (L. M. Benedict).

https://doi.org/10.1016/j.anbehav.2022.11.005
0003-3472/

memory ability often perform worse in tests designed to measure flexibility (Bebus et al., 2016; Croston et al., 2017; Hermer et al., 2021; Lewis & Kamil, 2006; Tello-Ramos et al., 2018). This raises questions of how animals may solve complex tasks in natural contexts when conditions change frequently.

The term 'behavioural flexibility' has been used to describe a variety of different behaviours (Audet & Lefebvre, 2017), but, traditionally in animal behaviour, cognitive flexibility is defined as the ability to rapidly relearn changing associations, allowing animals to better track a changing environment or learn the reversal rules (Badre & Wagner, 2006; Bond et al., 2007; Shettleworth, 2010; Strang & Sherry, 2014; Tello-Ramos et al., 2019). Behavioural flexibility is frequently measured in reversal associative learning tasks, either using a single reversal (e.g. Croston et al., 2017) or serial reversals (reviewed in Izquierdo et al., 2017). While both tasks require an animal to learn a reversed contingency, repeatedly reversing the same contingencies during the serial task may introduce different cognitive processes associated with flexibility. A single reversal task requires an individual to inhibit the recall of previously learned associations to associate new information with familiar cues during the reversal (Harlow, 1949; Strang & Sherry, 2014; Tello-Ramos et al., 2019). As there is only one reversed association, individuals can forget the first association to learn the second one. But this strategy poses a challenge for solving a serial reversal task because after the second reversal, animals need to relearn the first association, despite already learning to completely inhibit that association during the first reversal. Instead, to efficiently solve serial reversal tasks, an animal must learn to retain previous memories of both associations and learn the 'switching' or 'reversal rule' (Izquierdo et al., 2017).

Learning an abstract rule may integrate different parts of the brain and involve different cognitive processes than basic associative learning, such as executive functioning (Lai et al., 1995). Animals must first inhibit memories of previously learned associations (e.g. unlearning them) and then learn new associations (e.g. Lai et al., 1995). Therefore, one way to measure performance in a serial reversal task is to estimate how well animals learn and remember the reversal locations following each reversal, measured by the number of total errors before making a correct choice (e.g. Croston et al., 2017). Such errors made within a reversal learning context reflect the ability to switch between multiple memories without inhibition, potentially providing major advantages to individuals in variable and predictable environments.

Many changing environmental conditions occur across days, but it is still unclear to what extent animals can use time as a cue in associations or can incorporate time into learning to learn the predictable changes. Although it might be highly adaptive in variable environments to learn and use a rule to predict conditions across days or weeks, the 'stuck in time' hypothesis suggests that animals can only respond to stimuli in the present moment and cannot recall specific moments in the past to anticipate future events (Roberts, 2002; Tulving, 1985). Animals are well known to associate a stimulus or reward with internal cues generated by circadian rhythms, referred to as an individual's 'biological clock' (Mistlberger, 2009; Roberts & Feeney, 2009; Suddendorf & Corbalis, 2007). Animals can also anticipate daily rewards (e.g. Biebach et al., 1989), track short time intervals (e.g. Pavlov, 1927; reviewed in Shettleworth, 2010) and remember sequences of events after significant reinforcement (Devine et al., 1979; Roberts, 2002; Shimp, 1976; Shimp & Moffitt, 1974). But less clear are the mechanisms that may underlie the ability to learn to predict changes that occur across multiple days. Without this ability, the benefits of learning a rule may be limited to rules that can be associated with circadian rhythms (i.e. within a ca. 24 h period).

In this study, we designed an experiment to test whether wild food-caching mountain chickadees, *Poecile gambeli*, from two environments differing in environmental harshness and predictability (Croston et al., 2016, 2017; Heinen, Pitera, Sonnenberg, Benedict, Branch et al., 2021; Heinen, Pitera, Sonnenberg, Benedict, Bridge et al., 2021; Pitera et al., 2018; Tello-Ramos et al., 2018) can (1) learn the reversal rule in an eight-position spatial serial reversal learning task with daily location reversals in the wild and (2) learn to predict daily changes in the location of a food reward based on previous experience during the task. Both processes, learning the rules associated with repeated daily changes in a serial reversal task and learning to predict future events, are usually considered 'higher-level' cognitive processes, as they involve learning abstract strategies and rules that relate to learned associations (Shettleworth, 2010).

We conducted an eight-position spatial serial reversal task by alternating a rewarding food location each day between one of two locations in an eight-feeder spatial array (using 4 previously established feeder arrays; Fig. 1). This experimental set-up differed from classic serial reversal tasks in two key ways. First, instead of reversing the feeder locations after a learning criterion or fixed number of trials, rewarding feeder locations were switched every night, so every morning a different feeder provided food until the end of the day. This design allowed us to explore whether individuals could associate the switching rule with a daily change; birds were unlikely to be able to use circadian rhythms to solve this task because food was available at the same time every day (during daylight hours) but at a different location every other day. This design also allowed birds to continue to learn beyond a set learning criterion, but we could still assess performance at the end of each day to confirm that all birds both fully learned the new location and stopped visiting the previous location. Birds could forage ad libitum, and daily visits were grouped into 'trials'. A trial began when a bird visited a feeder at an array and ended when that bird found the food reward (after which chickadees typically left the array to eat or cache that seed individually; Table 1).

The other key difference between our task and a binary choice experiment was that our task presented eight equally likely locations for a food reward. This allowed us to measure 'total location

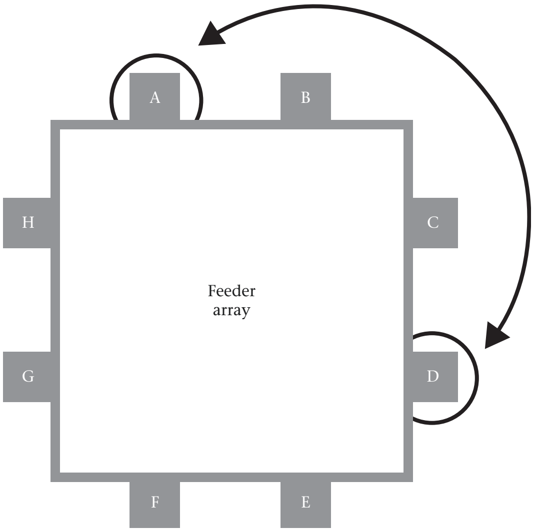

Figure 1. Example experimental set-up using a feeder array. Eight feeders (grey squares) were arranged on a square frame. For each bird, the rewarding feeder switched each day between two of the eight possible feeders (for example, feeder A and feeder D). Based on Croston et al. (2016). Individuals were not all assigned to the same feeders.

Table 1
Term definitions

Term	Definition
Trial	Starts when a bird visits a feeder at an array; ends when the bird visits the correct rewarding feeder
Reversal	When the rewarding feeder switches locations every day
Reversal feeder	The feeder that provided a food reward during the previous reversal but is no longer rewarding during the current reversal
Location errors	The number of feeders visited that did not provide a food reward (maximum of 7) per trial. The sum of neutral errors per trial and the reversal error (only one possible per trial)
Reversal errors	An incorrect visit to the reversal feeder; maximum of one per trial
Neutral errors	The number of feeders visited that never provided a food reward during the reversal task; maximum of six per trial
Performance	Generic term to describe the number of errors in a given context, e.g. location errors. Better performance describes fewer errors
Single reversal ability	Performance on the 5-day single reversal task before the first spatial serial reversal (Serial R1), used to assess cognitive flexibility before the serial reversal task
Reversal rule	The 'win-stay lose-shift' strategy: individuals learn that if food is not present in location A, they must shift to location D. Learning this rule perfectly should be reflected in 'one-trial, one-error' learning, in which individuals make one reversal error and no neutral errors in a given trial
Ability to predict	Individuals learn that the reversal rule is associated with changing days, and they make no errors in the first trial each day, going directly to the correct feeder

errors', or the number of nonrewarding feeders that were visited per trial (maximum of 7 total location errors per trial, out of 8 possible feeder locations; Table 1). Of these total location errors per trial, we could further differentiate whether birds made a 'reversal error' to the previously rewarding feeder during the previous day's reversal (maximum of 1 reversal error per trial; Strang & Sherry, 2014; Stanton et al., 2021) or a 'neutral error' to another feeder that did not provide a food reward during either reversal (maximum of 6 neutral errors per trial; see Seu et al., 2009 for a similar four-position single spatial reversal experiment in rats). This design differs from a binary choice reversal task, in which all errors are reversal errors by default, which specifically characterizes an individual's persistence in visiting the reversal location (one way to assess behavioural flexibility, reviewed in Izquierdo et al., 2017). But by differentiating between reversal and neutral errors, we could assess how birds learned the new, reversal association while simultaneously assessing whether birds remembered and relied on memories of the previous reversal location.

Neutral errors in a reversal learning context could provide information about imperfect learning and whether birds use a 'win-stay, lose-shift' strategy, or reversal rule, to solve the task. If birds are using this strategy (discussed in Strang & Sherry, 2014 using a binary choice task), then they should only make a single reversal error and no neutral errors during the first trial after each reversal, suggesting that an individual first makes an error to the previously rewarding choice and then immediately switches to the correct choice. When learning a serial reversal task, birds should demonstrate a continuum of learning from full inhibition of memories (i.e. unlearning the previous association) at the beginning of the serial reversal to perfect performance using a 'win-stay, lose-shift' strategy towards the end of the task, by fully retaining both memories of rewarded locations. Inhibition of previous memories associated with previous locations being unrewarded can be expected to result in a random search after each reversal, leading to neutral errors immediately following each reversal as an individual searches for the location that was rewarding 2 days prior. In contrast, completely uninhibited memories of both locations should be associated only with a single reversal error and no neutral errors, consistent with using the reversal rule. Between these two extremes, animals should go through intermediate stages of incomplete inhibition leading to imperfect memories, associated with moderate neutral errors after each reversal. Accordingly, neutral errors immediately after each reversal measured during the first trial should decrease throughout the task.

While binary choice tasks and other studies rely on reversal errors to measure cognitive flexibility (reviewed in Izquierdo et al., 2017), using neutral errors in addition to reversal errors provides additional detail to understand the resulting effect on learning and performance. Plus, in multiple-choice spatial learning tasks, each spatial location is unique and thus only one reversal error can be made per trial. In this case, using neutral errors provides greater detail to understand whether results are consistent with the effects of proactive interference and the influence of higher-level cognitive processes involving several brain regions including the hippocampus (e.g. Seu et al., 2009). Although this method does not differentiate between these mechanisms, it should capture the end result through variation in reversal learning. Ultimately, cognitive flexibility reflects how well and how quickly an animal learns the serial reversal task given all available choices, and thus measuring performance in a spatial serial reversal task using all potential errors provides the most ecologically relevant measure of flexibility.

We hypothesized that this task involves three types of learning in the following progression (1) learning the two reversing feeder locations (i.e. forming associations using spatial learning and memory) and maintaining memories for both locations despite one of them being unrewarded for a long period (full day); (2) learning the reversal rule associated with serial reversal (i.e. learning to switch between two rewarding feeders out of eight feeders using cognitive flexibility within a spatial context without making any neutral errors); and (3) learning to predict the next day's rewarding feeder location (i.e. learning to associate the serial reversal rule with time (i.e. days), so that individuals visit the new, correct location without first checking the previously rewarded location). We use the term 'performance' to describe the number of total location errors including both reversal and neutral errors birds made in different contexts. Good performance is reflected by a faster reduction in the number of location errors throughout the serial reversal task (i.e. days of the task). Optimal serial reversal performance is associated with a single reversal error on the first trial after each reversal, when an individual first visits the previously rewarded feeder and then immediately goes to the correct feeder that was rewarding 2 days prior without making any neutral errors. This behaviour could be associated with learning the reversal rule based on flexible memories.

We assessed (1) through performance on a spatial learning and memory task and a single reversal task before the serial reversal experiment began. We assessed (2) based on the number of location errors (both reversal and neutral) birds made during the first trial of each day, as each bird's first visit of the day should reveal where a bird expects to find food and does not represent relearning of the new location. Birds that perfectly learned the reversal rule could be expected to make only one error (reversal error) during each reversal or feeder switch (Strang & Sherry, 2014); they should go first to the previous day's rewarding feeder (i.e. the reversal feeder) but then immediately switch to the correct new location without making additional neutral errors. We additionally assessed reversal learning by testing for a reduction in the total number of reversal errors over the first 20 trials across the serial reversal task. Such a reduction is indicative of decreased persistence in making reversal errors throughout the task. We assessed (3) using the first trial of each day: if birds learned to predict the next feeder in the reversal sequence across days, they should go to the correct feeder without visiting the reversal feeder during the first visit of each day. If birds learned to predict the next location based on the reversal rule, then they should make no reversal errors, indicating that they go directly to the correct rewarding feeder during the first trial of each day. It is possible that performance measured by neutral errors is confounded by basic spatial learning and memory ability, so we additionally measured performance using only neutral errors without reversal errors. In addition, we directly tested whether spatial learning ability may indeed affect performance in our reversal task when measured either by all errors or just by neutral errors. We included individual performance on a spatial learning and memory task (which we have previously demonstrated to reflect learning ability; Branch et al., 2022; Sonnenberg et al., 2019) in all analyses.

We conducted this serial reversal task with wild mountain chickadees in their natural environment across two montane elevations with substantial differences in winter conditions. Winter conditions at higher elevations (ca. 2400 m) are consistently harsher (i.e. lower ambient temperatures, longer duration of snow cover) and less predictable (i.e. more frequent and unpredictable snowfall and more severe winter storms, causing unpredictable interruptions in food availability) compared to lower (ca. 1900 m) elevations (Heinen, Pitera, Sonnenberg, Benedict, Bridge et al., 2021; Kozlovsky et al., 2018; Pitera et al., 2018). Differences in daily foraging routines between low and high elevations during the winter were also consistent specifically with differences in environmental predictability (Pitera et al., 2018). If learning a rule is beneficial in more predictable environments, then we would expect birds from the milder, more predictable environment at low elevation to learn the alternating pattern better than birds from the harsher, less predictable environment at high elevation. Our predictions are based primarily on the possible advantage of learning a rule in predictable environments; however, there is also a difference in harshness across these elevations. We have previously found significant cognitive, morphological and behavioural differences between birds from each of these environments (Croston et al., 2016, 2017; Freas et al., 2012; Heinen, Pitera, Sonnenberg, Benedict, Bridge et al., 2021; Pitera et al., 2018; Tello-Ramos et al., 2018) and have found selection on spatial learning and memory at the high elevation (Sonnenberg et al., 2019). Because there is stronger selection on food-caching propensity and spatial learning and memory ability at higher elevations compared to lower elevations, we predicted that stronger proactive interference associated with stronger memories and higher memory load due to more food caching would negatively affect high-elevation birds' flexibility, reflected in their performance in learning a serial reversal task (Tello-Ramos et al., 2019). Similarly, we predicted that older birds would learn the reversal rule better than younger birds, according to previous work documenting differences in cognitive flexibility between these groups (Tello-Ramos et al., 2018).

METHODS

Study System

The study was conducted in winter 2020–2021 at Sagehen Experimental Forest in the Sierra Nevada (Sagehen Creek Field Station, University of California Berkeley, located 10 km north of Truckee, CA, U.S.A.) as part of our long-term mountain chickadee study (Croston et al., 2016, 2017; Freas et al., 2012; Kozlovsky et al., 2018; Tello-Ramos et al., 2018). During annual banding efforts (2014–2021), we trapped birds at nestboxes or feeders using mist nets and banded them with unique colour bands and a passive integrated transponder (PIT) tag with a unique alphanumeric ID (IB Technology, Leicestershire, U.K.). Age at initial capture was determined as 'juvenile' (less than 1 year of age) or 'adult' (at least 1 year or older) by using multiple plumage characteristics (Meigs et al., 1983; Pyle, 1997), breeding status and nestling banding records if available. Age in years was determined, when possible, from previous banding records (2013–2021).

Experimental Apparatus

Cognitive tasks in this study were conducted using spatial arrays of 'smart' feeders, established at the study system in 2014 and used annually for cognitive testing (Croston et al., 2016, 2017). Each of four arrays (two per elevation, ca. 1.2 km apart) consisted of eight feeders mounted equidistantly to a 122 × 122 cm aluminium frame and raised 3 m above the ground. Each feeder was equipped with radiofrequency identification (RFID) technology to detect PIT-tagged individuals and control access to food within the feeder via a mechanical door (Bridge & Bonter, 2011). The feeders could be set to three different 'modes': (1) 'open' mode, when the feeder doors were always open and food inside was clearly visible; (2) 'all' mode, when doors were closed until any PIT-tagged chickadee landed on the perch, triggering the door to open; (3) 'target' mode, when doors were closed and were programmed to only open for certain PIT-tagged birds. 'Target' mode was used during cognitive tests so that birds could be individually 'assigned' to a single rewarding feeder at one array. 'Open' and 'all' modes were used to habituate birds to the feeders. In all three modes, every visit by a PIT-tagged bird was recorded, whether or not the bird obtained a food reward.

Assessing Spatial Learning and Memory and Cognitive Flexibility

Before conducting the serial reversal experiment, birds were habituated to the feeders during 'open' and 'all' mode (20 November 2020–13 January 2021). To assess spatial learning and memory ability, birds were restricted to only one rewarding feeder in an array ('feeder A') for 5 days ('target' mode, 13–17 January 2021). Birds were assigned to a rewarding feeder pseudorandomly so that no individual was assigned to their most visited feeder during 'open' or 'all' mode and the new feeder was always on a different side of the square array. In a single spatial reversal task, birds were reassigned to a different rewarding feeder ('feeder D') on a different side of the same square array for 5 days ('target' mode, 17–20 January 2021). We use letters 'A' and 'D' to indicate that assigned feeders within a task were not adjacent, but the assigned feeder locations were not the same for every bird. Groups of birds previously assigned to the same feeder were reassigned to separate feeders to minimize social learning (Tello-Ramos et al., 2018).

Chickadees forage for single seeds, leaving the array to consume or cache the seed, so each visit to the rewarding feeder yields one food item. A 'trial' began when a bird first visited a feeder at an array and ended when it received a food reward; all feeder locations visited before finding the food reward were 'location errors' (Table 1). We used the mean number of location errors per trial in the first 20 trials as a metric of spatial learning and memory ability in the first task and of reversal learning ability in the second task (following our previous work: Croston et al., 2017; Heinen, Pitera, Sonnenberg, Benedict, Branch et al., 2021; Heinen, Pitera, Sonnenberg, Benedict, Bridge et al., 2021; Heinen, Benedict et al., 2021; Sonnenberg et al., 2019; Tello-Ramos et al., 2018). Our previous work showed that mean performance over the first 20 trials provides an ecologically relevant measure associated with differences in fitness.

Spatial Serial Reversal Tasks

We tested chickadee performance in two separate spatial serial reversal tasks in which the rewarding feeder alternated every day between two locations, switching at night when feeders were inactive so that each new day was another reversal. The first spatial serial reversal (Serial R1) immediately followed the spatial learning and memory and single reversal tasks and used the same feeder assignments assigned to individuals in the previous tasks (e.g. feeders A and D, 20–26 January 2021; Fig. 2). The second spatial serial reversal (Serial R2) started 16 days after the end of Serial R1 and used new feeder assignments for all birds (e.g. feeders E and H, 10–26 February 2021; Fig. 2). Feeders ‘E’ and ‘H’ were not adjacent to each other or to Serial R1 assignments.

We initially intended to conduct only one spatial serial reversal task over a longer interval to allow for more reversals and more time to detect significant trends. However, Serial R1 was interrupted by an unexpected snowstorm that compelled us to alter the operation of the feeders for the safety of the birds (Fig. 2). During the storm, feeders were set to the ‘open’ door mode. After the storm, feeders were reset to ‘all’ mode (5–10 February 2021, at low elevation; 8–10 February 2021, at high elevation) and then set to ‘target’ mode to start Serial R2.

The snowstorm interruption allowed us to design Serial R2 to test specifically whether experience with the two alternating locations affected serial reversal performance. During Serial R2, we reassigned each bird to new feeder locations (feeder E and H) to provide all birds with two new feeders to learn regardless of participation in the previous spatial serial reversal task. Thus, we were able to compare performance of (1) the same birds during Serial R1 and R2 and (2) between birds with and without previous experience in Serial R1. Furthermore, we did not repeat the spatial learning and memory task and single reversal task with the new feeder assignments. Instead, we started to alternate daily feeder assignments immediately, so that birds only had 1 day to learn each new feeder assignment before the second serial reversal task began (instead of 5 days, as in the previous cognitive tasks). This allowed us to assess the possible effect of giving birds less experience with the two feeder locations before beginning the serial reversal. Serial R2 then proceeded exactly the same as Serial R1 for the serial reversal but with more reversals (15 days).

Data were only included for each serial reversal if a bird participated in at least seven consecutive days of the task with at least 20 trials per day (see Appendix, Table A1, Fig. A1). Consecutive days were essential, because if a bird missed a day, it would miss an entire reversal and would experience a different feeder sequence compared to the other birds (e.g. instead of learning ‘ADADADA’, the bird that missed day 3 of Serial R1 might experience ‘ADDADA’). We used a minimum of 7 days for both serial reversal tasks because Serial R1 only lasted 7 days. We used a minimum of 20 trials, following our previous work (Croston et al., 2017; Heinen, Pitera, Sonnenberg, Benedict, Branch et al., 2021; Heinen, Pitera, Sonnenberg, Benedict, Bridge et al., 2021; Heinen, Benedict et al., 2021; Sonnenberg et al., 2019; Tello-Ramos et al., 2018), as the mean performance during the first 20 trials represents a meaningful measure associated with fitness (Sonnenberg et al., 2019). Using 20 trials per day also ensures individuals had minimum sufficient participation to learn the daily feeder location, since we did not use a learning criterion.

Assessing Serial Reversal Performance

In our system, a bird could make a maximum of seven location errors per trial, of which only one could be a reversal error and six could be neutral errors (Table 1). In general, when a bird visits any nonrewarding feeder (making either a reversal or neutral error) and does not find food there, it will visit other feeders until it finds the correct location and receives food. Thus, each bird receives one reward in every trial and leaves the feeder array until it begins another trial.

We expected the birds' behaviour to differ between the first reversal and subsequent reversals. After the first location switch (e.g. in the single reversal task), a bird should simultaneously learn the new association and inhibit its previously learned association. As it learns, it should be less likely to make a reversal error in each subsequent trial. But when a bird experiences multiple reversals (e.g. in the serial reversal task), it should learn not to inhibit the previously learned association because both associations are necessary to learn the reversal rule. This learning process should be

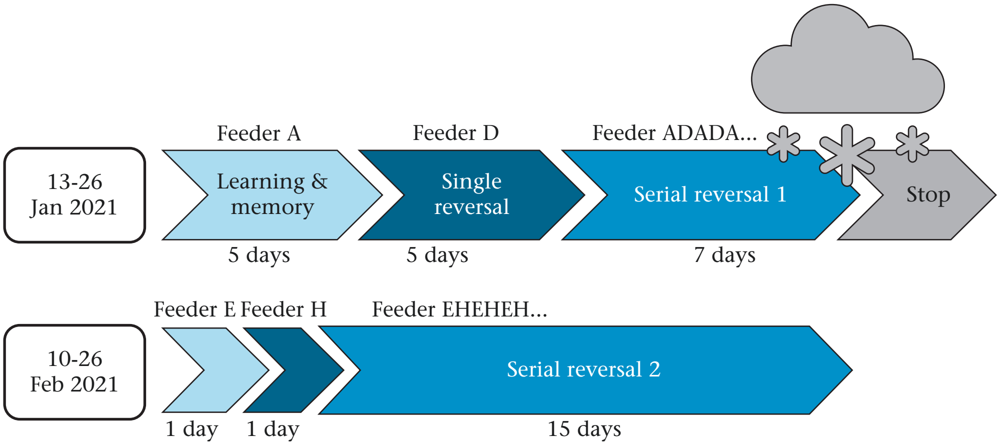

Figure 2. Timeline of experiments. During 13–26 January 2021, three cognitive tasks were conducted: (1) 5-day learning and memory task with birds assigned to feeder A; (2) 5-day single reversal task with birds assigned to feeder D; (3) 7-day serial reversal 1 task with feeder assignments alternating daily, ‘ADADA…’. Experiments stopped during a snowstorm. During 10–26 February 2021, birds were given (1) 1 day to learn feeder E, (2) 1 day to learn feeder H and (3) a 15-day serial reversal task 2 with feeder assignments alternating daily, ‘EHEHE…’.

reflected in performance across reversals, during the first trial after each reversal. In our serial reversal task, if the bird has ‘unlearned’ the previous association, it should treat the previous reversal location the same as the other nonrewarding feeders, making many errors as it searches for the correct feeder. As the bird learns not to inhibit the previous association, the bird should make fewer neutral errors during the first trial per day but may still make a reversal error, suggesting the bird can recall the location of both feeders. A bird is considered to have fully learned the serial reversal rule when it only makes one reversal error and makes no neutral errors in the first trial: the bird should visit the reversal feeder first, then consecutively visit the correct rewarding feeder (e.g. Strang & Sherry, 2014).

Metrics for Learning to Reverse

Location errors in trials 1–5 to measure performance

For both serial reversal tasks, we analysed the number of location errors each bird made per trial in the first five trials per day. We analysed the first trial separately from trials 2–5 per day because the first trial was entirely based on what birds learned prior to the current reversal. It provides specific information about where the birds expected to find food based on their previous experience when they visited the array for the first time following the night after each reversal (i.e. after being rewarded at a different feeder during the previous day). Performance in trial 1 per day does not represent relearning of the new location during that day. In contrast, during trials 2–5 per day, birds have already found the rewarding feeder during the first trial and thus could use that experience to find the correct feeder. We assessed trials 2–5 per day to test for consistency in performance following trial 1, and we expected to see evidence of learning across these trials because previous work has found variation in learning within five trials of similar cognitive tasks (Croston et al., 2017). Hereafter, we always refer to the trials ‘per day’ but may refer to these simply as ‘trial 1’ or ‘trials 2–5’.

Total location errors in the last five trials to assess daily learning

We also analysed the sum of location errors in the last five trials per day to assess whether individuals learned the association for that particular day. We used the sum of location errors in the statistical analyses instead of errors per trial to avoid complexities in fitting zero-inflated models, as we expected most trials towards the end of each day to have zero location errors. Performance at the end of each day was important to assess because we did not reverse associations based on individuals reaching a learning criterion, as is more common in the serial reversal literature. If a bird did not learn the feeder location by the end of a given day, we expected it to make more errors than expected by chance (i.e. more than 4.5 errors per trial, or more than 22.5 errors summed across five trials; Tille et al., 1996; assessed using one-sample Wilcoxon signed-rank tests). All birds in our study performed much better than expected by chance and, moreover, showed performance similar what is commonly used as a learning criterion.

We also used the sum of total location errors in the last five trials per day to explore whether birds demonstrated ‘anticipatory errors’, in which individuals make more errors just before the reversal switch. Anticipatory errors have been documented in single ‘within-session reversal’ tasks (i.e. when the reversal occurs once after a fixed number of trials; Rayburn-Reeves et al., 2011; Stagner et al., 2013). In our study, these errors might suggest birds use timing or counting to switch associations instead of local cues associated with the change in day, and we might see these errors increase throughout the serial reversal task despite improved performance in the first five trials per day.

Proportion of birds with perfect reversal performance

In addition, we analysed the proportion of birds that reached perfect serial reversal learning performance (i.e. only making a single error in trial 1, which is a reversal error) out of all the birds that participated each day and how this proportion changed throughout the task.

Metric for Learning to Predict

To assess whether birds learned to predict which feeder would provide food the next day, we tested whether or not birds made a reversal error in trial 1 per day (e.g. Rayburn-Reeves et al., 2011). If a bird learned to predict the alternating locations of the rewarding feeders, it should not make a reversal error in trial 1, instead going directly to the new rewarding feeder. Thus, the probability that a bird makes a reversal error in trial 1 should decrease across reversals (i.e. days) as the bird learns to predict the next feeder. This should coincide with an overall reduction in the number of location errors in trial 1 each day.

Metrics for Assessing Proactive Interference

Previous reversal trials

To assess the potential association between proactive interference and learning the serial reversal task, we used the total number of trials completed during the previous reversal (‘previous reversal trials’) to estimate total experience and potential learning of the reversal feeder location. If proactive interference negatively affects performance in the serial reversal, we expected birds that completed more previous reversal trials (and thus that formed stronger associations with the reversal feeder) would make more location errors. Stronger associations with a location should lead to more inhibition of previous memories and hence more neutral errors would be expected.

Sum of reversal errors in the first 20 trials

We also analysed the sum of reversal errors made in the first 20 trials per day (maximum 20 reversal errors; maximum of one per trial) to assess ‘reversal’ persistence, when a bird continued to visit the feeder that was rewarded during the previous day even after it located the new rewarding feeder in trial 1. As birds learn the reversal rule and learn to overcome proactive interference, we should see a reduction in the number of reversal errors across the serial reversal task. If birds at high elevation have more proactive interference than birds at low elevation, we would expect low-elevation birds to reduce the number of reversal errors faster (Croston et al., 2017). For birds that participated in both serial reversal tasks, we also analysed the number of reversal errors made in Serial R2 to previously rewarding feeders from Serial R1.

Comparing performance between serial reversal tasks

We compared performance (location errors in trial 1) between the two serial reversal tasks by (1) comparing birds that participated in both tasks to birds that participated in only one task and (2) comparing performance between tasks within birds that participated in both Serial R1 and R2.

Statistical Analyses

Data from Serial R1 and R2 were analysed separately due to the different experimental procedures (i.e. the amount of time to learn the two rewarding feeder locations) and possible effects from conducting experiments sequentially. We primarily used linear regression models fitted using R version 4.1.0 (R Core Team, 2018). Before and during model fitting, we checked all assumptions of linear models, using the R package 'DHARMa' (Hartig, 2020) to check Q–Q plots, residual by predicted value plots and goodness-of-fit tests. Generalized linear mixed-effects models (GLMM) were fitted for daily data that contained repeated measures (individuals repeated each day). The conclusions of Nakagawa and Cuthill (2007) and Schielzeth et al. (2020) were considered during model fitting and reporting.

Learning the locations

The two rewarding feeder locations used for Serial R1 were the same as those learned by birds during the 5-day spatial learning and memory task and the 5-day single reversal task (Fig. 2). The mean number of location errors per trial over the first 20 trials of each task (not per day) were used to assess performance on each task. These metrics were used as predictor variables to assess learning and memory ability and single reversal ability, respectively, in serial reversal analyses. Separate models were fitted for each cognitive ability to avoid possible covariance, due to the hypothesized trade-off between spatial learning and memory and cognitive flexibility (Tello-Ramos et al., 2018, 2019). The two new locations for Serial R2 were introduced 2 days before the start of Serial R2 (Fig. 2), were available for 1 day each and were also evaluated using the mean number of location errors in the first 20 trials. However, we did not use these measures as predictor variables in our models, as these locations were the third and fourth learned locations for birds that had already participated in Serial R1.

For all four feeder locations, we used linear regressions to evaluate how mean performance varied across elevations (high and low) and age groups (juvenile and adult). We used one-sample Wilcoxon signed-rank tests to compare the mean performance for each learned location to chance level (given eight choices, random expectation is 4.5 errors; Tille et al., 1996).

Learning to reverse

We fitted four Poisson generalized linear mixed-effects models (GLMM) with the number of location errors per bird per day (1) in trial 1, (2) trials 2–5, (3) the sum of the last five trials and (4) in trials 2–5 before making a reversal error. All models included a random intercept of individual bird ID to adjust for repeated measures. All models included the same predictor variables: day of the experiment (i.e. reversal number), elevation (high or low), cognitive task performance (either spatial learning and memory ability from the first 5-day task or single reversal performance from the second 5-day task; Fig. 2), either age class (juvenile or adult) or age in years (0–8 years old) and interaction effects of day∗elevation and cognitive performance∗elevation. We also included the number of previous reversal trials to assess the effect of total experience from the previous reversal, which should be indicative of the strength of the reversal association; more trials can be expected to result in stronger associations. We first fitted models containing all predictor variables and dropped nonsignificant interactions and variables from reported analyses (except for the main day∗elevation interaction, which was always included). Numeric variables were centred and scaled, and the previous reversal trials were log-transformed. Models using data from trials 2–5 also included trial number as a categorical variable to compare performance between trials. We did not include trial number as a numeric variable because we did not expect a linear relationship between performance (i.e. learning) and trial number. Models using data from the last five trials per day used the total number of trials completed that day, instead of previous reversal trials, to control for total learning during that reversal. Models using the number of neutral errors made before making a reversal error during trials 2–5 excluded trials in which birds did not make a reversal error at all.

To assess previous reversal trials in Serial R1, data from the first reversal (day 1) had to be excluded because there were no data from a previous reversal. To assess whether this exclusion biased the data, the models were also fitted using all 7 days of data from Serial R1 without previous reversal trials (Appendix, Table A2). There was no difference in the main results.

We also fitted a Gaussian GLMM using the proportion of birds each day that made exactly one reversal error in trial 1, with no neutral errors. The model included a fixed effect of day, elevation and a day∗elevation interaction. The model was weighted by the number of total birds per day.

Learning to predict

We created a binary categorical response variable indicating whether or not a bird made a reversal error in trial 1 per day. We fitted logistic GLMMs using a logit link with a random intercept of individual and fixed effects of day, elevation, age, cognitive ability and the number of previous reversal trials. Learning to predict which feeder would be rewarding each day should be associated with a significantly lower chance of visiting the reversal feeder in trial 1 across the serial reversal task.

Possible effects of proactive interference

In addition to analysing the effect of the number of previous reversal trials, we also used GLMMs to compare performance between birds without experience in Serial R1 with birds that participated in both Serial R1 and R2, to assess performance between the two serial reversal tasks and to test for elevational differences in the sum of reversal errors for the first 20 trials across the serial reversal task.

To compare performance of birds that participated in both serial reversal tasks with those that only participated in Serial R2, we fitted a Poisson GLMM model using the number of location errors from trial 1 per day of Serial R2. The model included a random intercept of individual and fixed effects of Serial R1 participation status (categorical), day, elevation, previous reversal trials and interactions of day∗elevation and participation∗elevation.

To explore differences in performance between the two serial reversal tasks within the same birds, we fitted a Poisson GLMM model using the number of location errors made in trial 1 per day of Serial R1 and R2 using only birds that completed both tasks. The model included a random intercept of individual and fixed effects of day, elevation, previous reversal trials, task (Serial R1 or R2) and interactions of day∗elevation and task∗elevation.

To assess the sum of reversal errors in the first 20 trials per day of Serial R1 and R2, we fitted three GLMM models with a random intercept of individual and fixed effects of day, elevation, number of previous reversal trials and an interaction of day∗elevation. One model used reversal errors made in Serial R1, one model used reversal errors made in Serial R2, and the third model used errors made in Serial R2 to the feeders assigned during Serial R1. We thus limited the Serial R2 data set to birds that participated in both serial reversal tasks.

Statistical software

All models were fitted using 'lme4' package (Bates et al., 2015) in R version 4.1.0 (R Core Team, 2018) and evaluated using Wald chi-square, and P values were calculated from ANOVA tests using the 'car' package. Tukey adjusted post hoc pairwise comparisons were calculated using 'emmeans' (Lenth, 2020). Residuals were analysed using 'DHARMa' (Hartig, 2020), and modified R^2 values were calculated according to Nakagawa and Schielzeth (2013) using the 'performance' package (Ludecke et al., 2020). Seven optimizers were evaluated using 'allFit' from the 'lme4' package for each model before selecting the 'bobyqa' optimizer from the 'lme4' package.

The number of trials completed on either the previous day or the current day was log-transformed due to a strong right skew. Numeric variables were centred and scaled. Figures were created using ‘ggplot2’ (Wickham, 2016), ‘ggeffects’ (Ludecke, 2018) and ‘raincloud’ plots (Allen et al., 2021).

Ethical Note

To the best of our knowledge, no birds were harmed by the collection of these data and birds were only handled for a few minutes during banding. To band adult and juvenile birds, we placed up to three plastic colour bands around the birds’ legs in a process that took no more than 30 s. One of these colour bands contained a PIT tag. If the bird had been previously banded with a metal numeric band issued by the U.S. Geological Survey Bird Banding Laboratory, only two colour bands were added and the colour band containing the PIT tag was placed on the opposite leg as the metal band. We detected no negative effects of using PIT tags and colour bands during our study. After bands were attached, birds were visually inspected for age and sex and were measured for wing length using flattened wing length. Total handling time was only a few minutes and birds were released immediately. To reduce stress during handling, banding efforts were not conducted during adverse conditions (i.e. wind, precipitation), birds were banded as soon as possible after capture and total handling time was minimized further during cold temperatures (ca. <5 °C). The study was approved by the University of Nevada Reno Institutional Animal Care and Use Committee (Protocol numbers 00818, 00046 and 00603) and was in accordance with California Department of Fish and Wildlife Permit D-0011776516-4. For banding efforts, we followed U.S. Federal Bird Banding Permit 22878.

RESULTS

We first present results for Serial R1, then for Serial R2, and finally compare both serial reversal tasks to assess evidence for proactive interference. We present results organized by predictions for each serial reversal task: learning the locations, learning to reverse and learning to predict. Age was not a significant predictor in any GLMM and was removed from reported results (Appendix, Tables A3, A5).

Serial R1: Learning the Locations

In the 5-day spatial learning and memory task, birds performed better than chance (one-sample Wilcoxon signed-rank tests: $V = 0$, $P < 0.001$; mu = 4.5 location errors per trial), making an average ± SE of 1.07 ± 0.38 mean location errors per trial in the first 20 trials. Performance did not vary significantly between elevations or age groups (linear regression: elevation: $F_{1,171} = 0.93$, $P = 0.34$; age: $F_{1,171} = 2.77$, $P = 0.10$; $R^2_{adj} = 0.01$; $N = 97$).

In the 5-day single reversal task, birds performed better than chance (one-sample Wilcoxon signed-rank tests: $V = 0$, $P < 0.001$; mu = 4.5 location errors per trial), making an average ± SE of 0.50 ± 0.24 mean location errors per trial in the first 20 trials. Single reversal performance also did not vary significantly between elevations or age groups (linear regression: elevation: $F_{1,171} = 2.56$, $P = 0.11$; age: $F_{1,171} = 0.01$, $P = 0.91$; $R^2_{adj} = 0.004$; $N = 97$).

Serial R1: Learning to Reverse

Serial R1, trial 1

During Serial R1, chickadees improved their trial 1 performance, making fewer location errors in trial 1 each subsequent day of the task. Birds at low elevation improved significantly faster than birds at high elevation ($N = 97$; Table 2, Fig. 3a). There was no significant effect of the number of previous reversal trials or spatial learning and memory ability on trial 1 performance, but birds with better single reversal ability performed significantly better in trial 1 throughout Serial R1 (Table 2).

There were significant interactions between elevation and day and elevation and single reversal ability (Table 2, Appendix, Fig. A2). Post hoc GLMM analyses indicated that the main effects in the model were driven by birds at low elevation: birds' performance in trial 1 improved significantly with days and birds with better single reversal ability also performed better in trial 1 (Poisson GLMM for low elevation: day: estimate = −0.18, $\chi^2_1 = 15.82$, $P < 0.001$, $N = 46$; single reversal ability: estimate = 0.30, $\chi^2_1 = 30.96$, $P < 0.001$; number of trials previous day: $\chi^2_1 = 0.38$, $P = 0.54$, $R^2c = 0.27$, $R^2m = 0.19$; $N = 46$). In contrast, at high elevation, there was no significant effect of day or single reversal ability on performance in trial 1 (Poisson GLMM for high elevation: day: $\chi^2_1 = 0.45$, $P = 0.50$; single reversal ability: $\chi^2_1 = 0.61$, $P = 0.44$; number of trials previous day: $\chi^2_1 = 0.73$, $P = 0.39$; $R^2c = 0.13$, $R^2m = 0.01$; $N = 51$).

Birds at both elevations overall performed better than chance in trial 1, on average making only a few location errors by the last day of Serial R1 (mean ± SE location errors in trial 1 on day 7 at low elevation: 1.33 ± 1.32; one-sample Wilcoxon signed-rank test: $V = 577.5$, $N = 46$, $P < 0.001$; mean ± SE location errors in trial 1 on day 7 at high elevation: 2.25 ± 1.49; one-sample Wilcoxon signed-rank test: $V = 2434.5$, $N = 51$, $P < 0.001$).

In Serial R1, the proportion of birds that made one reversal error and no neutral errors in trial 1 was greater at low elevation compared to high elevation and significantly increased at both elevations across days of the task (linear regression: elevation: $F_{1,11} = 7.28$, $P = 0.02$; day: estimate ± SE = 0.03 ± 0.01, $F_{1,11} = 12.20$, $P = 0.005$; $R^2_{adj} = 0.57$; $N = 97$; Fig. 4a).

Serial R1, trials 2–5

After finding the correct feeder in trial 1, chickadees dramatically reduced the number of location errors they made in subsequent trials each day, making an average ± SE of 0.23 ± 0.64 errors per trial across trials 2–5 in Serial R1 (Fig. 3a). In contrast to trial 1, performance on trials 2–5 did not vary significantly across days (Fig. 3a) or between elevations. Trial 2–5 location errors were also not significantly associated with spatial learning and memory ability; however, birds that completed more previous reversal trials performed better during trials 2–5 (Table 2).

There was no overall effect of single reversal ability on trial 2–5 performance in Serial R1, but there was a significant interaction between elevation and single reversal ability (Table 2). Post hoc GLMM indicated that, at high elevation, birds with better single reversal ability also performed better during trials 2–5 (Poisson GLMM: day: $\chi^2_1 = 0.83$, $P = 0.36$; single reversal ability: estimate = 0.24, $\chi^2_1 = 11.04$, $P < 0.001$; number of previous reversal trials: $\chi^2_1 = 3.36$, $P = 0.07$; trial number: $\chi^2_3 = 3.48$, $P = 0.32$; $R^2c = 0.12$, $R^2m = 0.06$). In contrast, there was no significant effect of single reversal ability on trials 2–5 at low elevation (Poisson GLMM: day: $\chi^2_1 = 0.002$, $P = 0.97$; single reversal ability: $\chi^2_1 = 1.67$, $P = 0.20$; number of previous reversal trials: $\chi^2_1 = 1.63$, $P = 0.20$; trial number: $\chi^2_3 = 15.60$, $P = 0.001$; $R^2c = 0.14$, $R^2m = 0.04$).

Serial R1, last five trials

The sum of location errors made in the last five trials per day was overall significantly better than chance (one-sample Wilcoxon signed-rank test: $V = 0$, $N = 97$, $P < 0.001$) and did not vary significantly across days (Table 3, Appendix, Fig. A3). Birds at low elevation made significantly fewer location errors in the last five trials (average ± SE = 0.38 ± 0.736 total location errors in last five

Table 2
Serial R1 GLMM model results with location errors per trial during trials 1–5

Focal variable	Effect	Spatial Serial Reversal 1 ($N = 97$)							
		Trial 1				Trials 2–5			
		Estimate	SE	χ^2	*P*	Estimate	SE	χ^2	*P*
Spatial learning and memory	Day	−0.18	0.04	16.52	**<0.001**	−0.006	0.06	0.01	0.92
	Elevation	0.32	0.08	15.17	**<0.001**	—	0.13	0.17	0.68
	Spatial learning and memory	0.07	0.04	3.38	0.07	0.05	0.06	0.62	0.43
	Previous reversal trials	<0.001	0.07	0.00	0.99	−0.25	0.1	6.22	**0.01**
	Trial	—	—	—	—	—	—	11.83[a]	**0.008**
	Day∗elevation	0.17	0.06	8.05	**0.005**	−0.05	0.09	0.34	0.56
	Bird ID (random intercept)	—	—	—	$\sigma^2 = 0.084$	—	—	—	$\sigma^2 = 0.21$
				R^2c = 0.25, R^2m = 0.09				R^2c = 0.13, R^2m = 0.02	
				AIC = 2034.1				AIC = 2989.7	
Single reversal ability	Day	−0.18	0.04	15.81	**<0.001**	−0.005	0.06	0.01	0.94
	Elevation	0.32	0.07	18.76	**<0.001**	0.01	0.12	0.01	0.92
	Single reversal ability	0.25	0.04	31.96	**<0.001**	−0.11	0.07	2.07	0.15
	Previous reversal trials	0.07	0.06	1.10	0.29	−0.22	0.1	5.07	**0.02**
	Trial	—	—	—	—	—	—	11.77[a]	**0.008**
	Day∗elevation	0.15	0.05	6.68	**0.01**	−0.06	0.09	0.41	0.52
	Single reversal ability∗elevation	−0.19	0.08	6.26	**0.01**	0.44	0.13	11.76	<0.001
	Bird ID (random intercept)	—	—	—	$\sigma^2 = 0.05$	—	—	—	$\sigma^2 = 0.17$
				R^2c = 0.26, R^2m = 0.17				R^2c = 0.1, R^2m = 0.04	
				AIC = 2010.8				AIC = 2981.1	

Separate models were fitted for spatial learning and memory and single reversal ability and trial 1 and trials 2–5. Conditional and marginal R^2 (R^2c and R^2m) were calculated according to Nakagawa and Schielzeth (2013). Unstandardized regression coefficients used as estimates. Estimates for elevation are relative to high elevation. All χ^2 statistics use 1 degree of freedom, unless otherwise indicated. Bold text indicates *P* values less than the significance level ($\alpha = 0.05$).

[a] The χ^2 statistic reports 3 degrees of freedom.

trials , $N = 46$) than birds at high elevation (0.52 ± 0.77 total location errors in last five trials, $N = 51$) and their overall performance was consistent with better than 90% criterion (<0.5 errors/5 trials). Furthermore, birds completing more trials on the same day made fewer location errors in the last five trials; thus, birds with more experience with the daily feeder location had better performance (Table 3).

Serial R1: Learning to Predict

Birds, on average, were significantly less likely to make a reversal error in trial 1 as Serial R1 progressed (Table 4, Fig. 3b). Spatial learning and memory ability was not a significant predictor, but birds that completed fewer previous reversal trials and birds

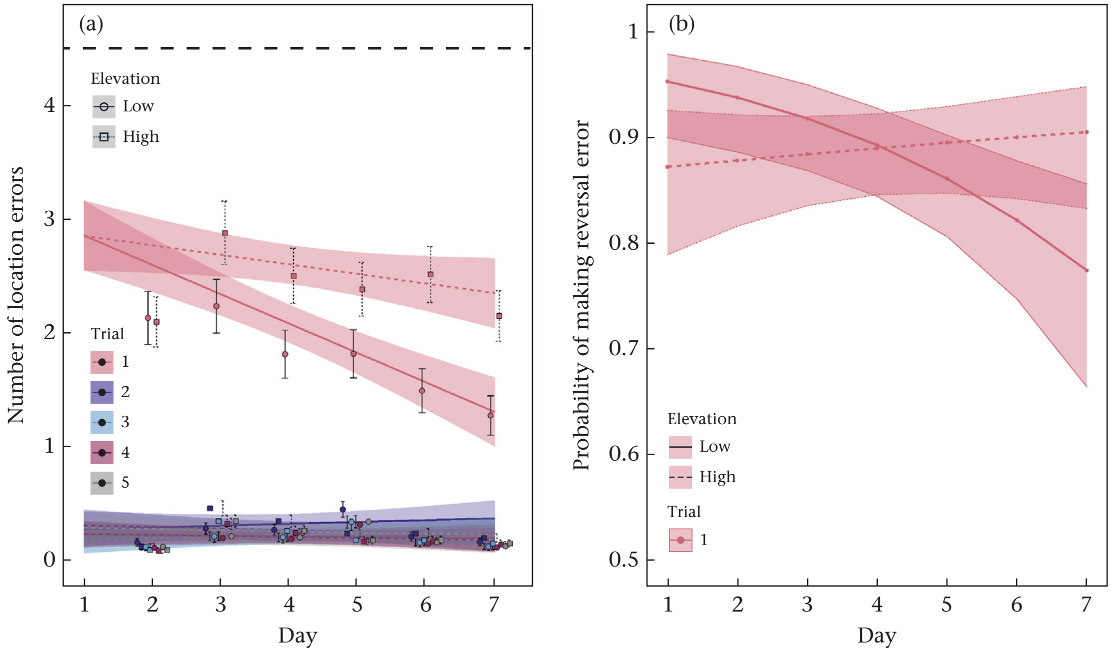

Figure 3. For Serial R1, (a) the number of location errors per trial for trials 1–5 per day and (b) the predicted probability of visiting the previous feeder (i.e. make a reversal error) in trial 1 per day. Shaded areas indicate 95% confidence intervals generated using the 'ggplot lm' method. In (a), points and error lines indicate estimated marginal means and standard errors were calculated from reported GLMMs. Heavy dashed lines indicate expected random error values. In (b), the predicted probability was calculated using reported GLMMs and 'ggeffects'.

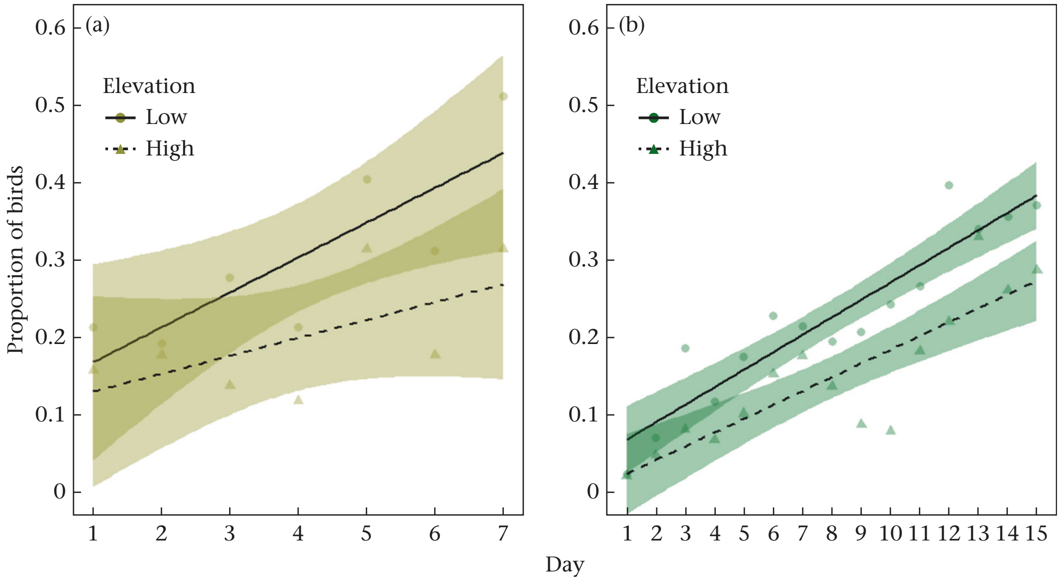

Figure 4. Proportion of birds that showed 'one-trial, one-error' learning (one reversal error, no neutral errors) during trial 1 of (a) Serial R1 and (b) Serial R2. Lines represent linear regression and shaded areas indicate standard error of the regression line.

Table 3
Model results for the total errors in the last five trials in Serial R1 and R2

Effect	Serial R1 ($N = 97$)				Serial R2 ($N = 235$)			
	Estimate	SE	χ^2_1	P	Estimate	SE	χ^2_1	P
Day	−0.04	0.09	0.24	0.63	−0.17	0.03	24.33	**<0.001**
Elevation	0.35	0.14	6.32	**0.01**	−0.15	0.09	2.66	0.10
Total daily trials	−0.42	0.11	14.43	**<0.001**	−0.37	0.06	40.11	**<0.001**
Day∗elevation	−0.08	0.11	0.59	0.44	0.01	0.06	0.01	0.90
Bird ID (random intercept)	—	—	—	$\sigma^2 = 0.13$	—	—	—	$\sigma^2 = 0.22$
			R^2c = 0.15, R^2m = 0.06				R^2c = 0.20, R^2m = 0.05	

Conditional and marginal R^2 (R^2c and R^2m) were calculated according to Nakagawa and Schielzeth (2013). Unstandardized regression coefficients used as estimates. Elevation estimates are relative to high elevation in Serial R1 and low elevation in Serial R2. Bold text indicates P values less than the significance level ($\alpha = 0.05$).

with better single reversal ability were less likely make a reversal error (Table 4).

There was a significant interaction between day and elevation, and post hoc GLMM models showed that the likelihood of making a reversal error only decreased significantly at low elevation, but not at high elevation. On the last day of Serial R1, ca. 21% of birds at low elevation did not make a reversal error in trial 1 (Fig. 5a). Moreover, when these birds made no reversal errors, they also made zero or one neutral error (Fig. 5b), suggesting that when birds did not visit the previous reversal feeder, they went directly to the correct rewarding feeder instead.

There was also a significant interaction between single reversal ability and elevation (Table 4), although the model reported a singular fit and thus interpretation may be limited to chickadees in this sample (see Appendix for further details). Post hoc GLMM models indicated that the effect of single reversal ability was also only present at low elevation, and there was no significant effect of the number of previous reversal trials (logistic GLMM: day: estimate = −0.05, $\chi^2_1 = 7.43$, $P = 0.006$; single reversal ability: estimate = 1.13, $\chi^2_1 = 15.66$, $P < 0.001$; previous reversal trials: $\chi^2_1 = 3.39$, $P = 0.07$; R^2c = NA, R^2m = 0.33). In contrast, at high elevation, the probability of making a reversal error did not change significantly over time and was not significantly associated with single reversal ability, but more previous reversal trials were associated with a higher probability of making a reversal error (logistic GLMM: day: $\chi^2_1 = 0.87$, $P = 0.35$; single reversal ability: $\chi^2_1 = 3.57$, $P = 0.06$; previous reversal trials: estimate = 0.84, $\chi^2_1 = 5.28$, $P = 0.02$; R^2c = NA, R^2m = 0.12). Both post hoc models were overfitted and R^2c could not be calculated (reported as R^2c = NA). However, we had to include individual ID as a random effect due to pseudoreplication issues.

Serial R2: Learning the Locations

Birds had 1 day to learn the first feeder location of Serial R2 (i.e. feeder E) before Serial R2 began. While learning this location, birds performed better than chance in the first 20 trials (one-sample Wilcoxon signed-rank test: $V = 0$, $N = 235$, $P < 0.001$), making an average ± SE of 0.68 ± 0.31 mean location errors per trial in the first 20 trials. Juveniles overall performed worse than adults, and there was a significant interaction between elevation and age: juveniles at high elevation performed better than juveniles at low elevation, but there was no difference between elevations for adults (linear regression: elevation: $F_{1,231} = 0.72$, $P = 0.40$; age (categorical): $F_{1,231} = 26.60$, $P < 0.001$; age∗elevation: $F_{1,231} = 4.81$, $P = 0.03$; $R^2_{adj} = 0.12$; $N = 235$).

Birds next had 1 day to learn the second feeder location (i.e. feeder H, first reversal) before Serial R2 began. Birds also performed better than chance in the first 20 trials (one-sample Wilcoxon signed-rank test: $V = 0$, $N = 235$, $P < 0.001$), making an average ± SE of 0.65 ± 0.36 mean location errors per trial in the first 20 trials. While learning the location of feeder H, juveniles overall

Table 4
Logistic GLMM results for the odds of making a reversal error during trial 1 each day of Serial R1 and Serial R2

Focal variable	Effect	Spatial Serial Reversal 1 ($N = 97$)				Spatial Serial Reversal 2 ($N = 180$, $N = 164$)			
		Estimate	SE	χ^2_1	P	Estimate	SE	χ^2_1	P
Spatial learning and memory	Day	−0.51	0.19	7.28	**0.007**	−0.03	0.1	0.09	0.77
	Elevation	0.01	0.27	0.002	0.97	−0.29	0.13	4.70	**0.03**
	Spatial learning and memory	0.1	0.13	0.58	0.45	−0.15	0.06	7.07	**0.008**
	Previous reversal trials	0.65	0.25	6.82	**0.009**	0.39	0.12	10.01	**0.002**
	Day∗elevation	0.73	0.27	7.33	**0.007**	0.16	0.12	1.68	0.19
	Bird ID (random intercept)	—	—	—	$\sigma^2 = 0.004$	—	—	—	$\sigma^2 = 0.05$
				R^2c = 0.09, R^2m = 0.09				R^2c = 0.05, R^2m = 0.03	
				AIC = 444.6				AIC = 2079.9	
Single reversal ability	Day	−0.54	0.2	7.38	**0.007**	−0.02	0.11	0.03	0.86
	Elevation	−0.32	0.32	0.97	0.32	−0.22	0.14	2.41	0.12
	Single reversal ability	0.93	0.23	16.38	**<0.001**	0.03	0.06	0.17	0.68
	Previous reversal trials	0.79	0.27	8.66	**0.003**	0.34	0.13	7.30	**0.007**
	Day∗elevation	0.73	0.28	6.97	**0.008**	0.15	0.13	1.37	0.24
	Single reversal ability∗elevation	−1.35	0.31	18.58	**<0.001**	NS			
	Bird ID (random intercept)	—	—	—	$\sigma^2 = 0.004$	—	—	—	$\sigma^2 = 0.08$
				Singular fit				R^2c = 0.04, R^2m = 0.02	
				AIC = 420.9				AIC = 1928.3	

Separate models were fitted for spatial learning and memory and single reversal ability. Conditional and marginal R^2 (R^2c and R^2m) were calculated according to Nakagawa and Schielzeth (2013). ‘NS’ indicates nonsignificant term removed from final model. Estimates on the log-odds scale. Elevation estimates relative to high elevation. Models fitted with spatial learning and memory for Serial R2 had a sample size of 180; models fitted with single reversal ability for Serial R2 had a sample size of 164. Bold text indicates P values less than the significance level ($\alpha = 0.05$).

performed significantly worse than adults, but there was no significant difference in performance between elevations and no interaction effect between elevation and age (linear regression: elevation: $F_{1,231} = 0.002$, $P = 0.96$; age (categorical): $F_{1,231} = 10.97$, $P = 0.001$; $R^2_{adj} = 0.04$; $N = 235$).

Serial R2: Learning to Reverse

Serial R2, trial 1

Performance in trial 1 improved significantly across days in Serial R2 (Table 5, Appendix, Fig. 6a). There was no overall effect of elevation, but there was a significant interaction between day and elevation: birds at low elevation improved their performance at a higher rate and made significantly fewer location errors than birds at high elevation by the end of Serial R2. Birds that completed more previous reversal trials performed significantly worse in trial 1, but there was no effect of spatial learning and memory ability or single reversal ability (Table 4).

In Serial R2, the proportion of birds that made one reversal error and no neutral errors in trial 1 was significantly greater at low elevation compared to high elevation and significantly increased at both elevations across days of the task (linear regression: elevation: $F_{1,27} = 18.12$, $P < 0.001$; day: estimate ± SE $= 0.02 \pm 0.002$, $F_{1,27} = 84.09$, $P < 0.001$; $R^2_{adj} = 0.77$; $N = 235$; Fig. 4).

Serial R2, trials 2–5

After locating the rewarding feeder during trial 1 each day of Serial R2, chickadees again showed almost perfect performance in

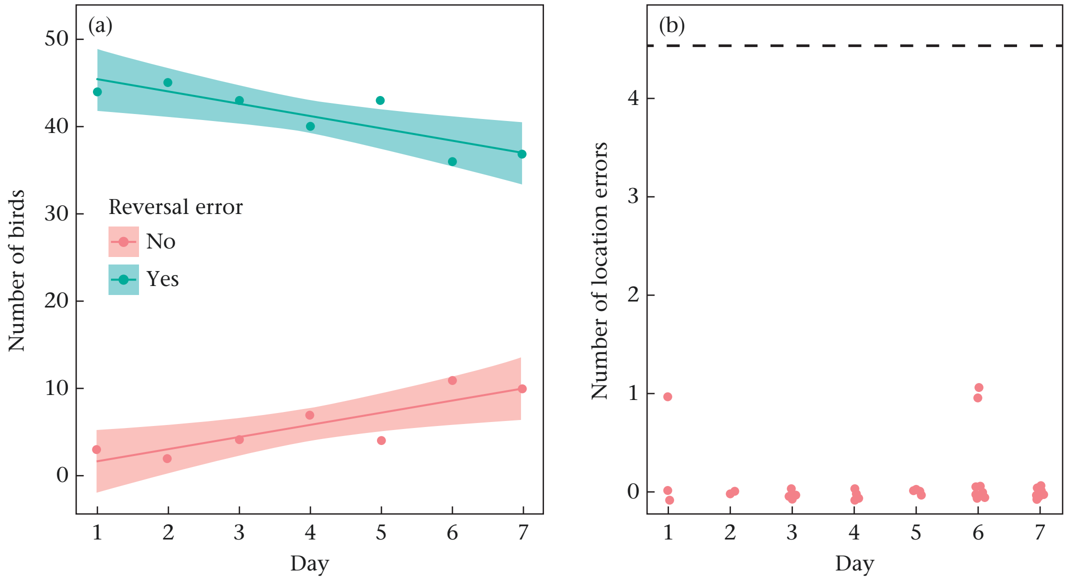

Figure 5. At low elevation in Serial R1: (a) the number of birds per day that either made or did not make a reversal error during trial 1 after each reversal and (b) the number of neutral errors made by birds that did not make a reversal error during trial 1.

Table 5
Serial R2 model results for the number of location errors in trials 1–5

Focal variable	Effect	Spatial Serial Reversal 2							
		Trial 1				Trials 2–5			
		Estimate	SE	χ^2	P	Estimate	SE	χ^2	P
Spatial learning and memory									
($N = 180$)	Day	−0.23	0.02	92.66	**<0.001**	−0.59	0.04	216.45	**<0.001**
	Elevation	0.03	0.04	0.67	0.41	0.29	0.09	10.37	**0.001**
	Spatial learning and memory	−0.01	0.03	0.34	0.56	0.07	0.04	2.68	0.10
	Previous reversal trials	0.1	0.02	11.53	**<0.001**	0.22	0.05	18.36	**<0.001**
	Trial	—	—	—	—	—	—	16.90[a]	**<0.001**
	Day∗elevation	0.07	0.03	5.70	**0.02**	0.36	0.05	54.68	**<0.001**
	Bird ID (random intercept)	—	—	—	$\sigma^2 = 0.02$	—	—	—	$\sigma^2 = 0.22$
				R^2c = 0.16, R^2m = 0.09				R^2c = 0.20, R^2m = 0.09	
				AIC = 8473.2				AIC = 12914.9	
Single reversal ability									
($N = 164$)	Day	−0.23	0.03	81.33	**<0.001**	−0.61	0.04	190.07	**<0.001**
	Elevation	0.04	0.04	0.91	0.34	0.29	0.1	8.72	**0.003**
	Single reversal ability	0.02	0.03	0.80	0.37	−0.01	0.04	0.08	0.78
	Previous reversal trials	0.1	0.02	9.22	**0.002**	0.22	0.05	17.37	**<0.001**
	Trial	—	—	—	—	—	—	15.19[a]	**0.002**
	Day∗elevation	0.07	0.03	5.32	**0.02**	0.38	0.05	54.80	**<0.001**
	Bird ID (random intercept)	—	—	—	$\sigma^2 = 0.03$	—	—	—	$\sigma^2 = 0.22$
		`		R^2c = 0.16, R^2m = 0.09				R^2c = 0.20, R^2m = 0.09	
				AIC = 7753.5				AIC = 11777.9	

Separate models were fitted for spatial learning and memory and single reversal ability, due to covariance. Conditional and marginal R^2 (R^2c and R^2m) were calculated according to Nakagawa and Schielzeth (2013). Unstandardized regression coefficients used as estimates. Elevation estimates are relative to high elevation. All χ^2 statistics use 1 degree of freedom, unless otherwise indicated. Bold text indicates P values less than the significance level ($\alpha = 0.05$).

[a] The χ^2 statistic reports 3 degrees of freedom.

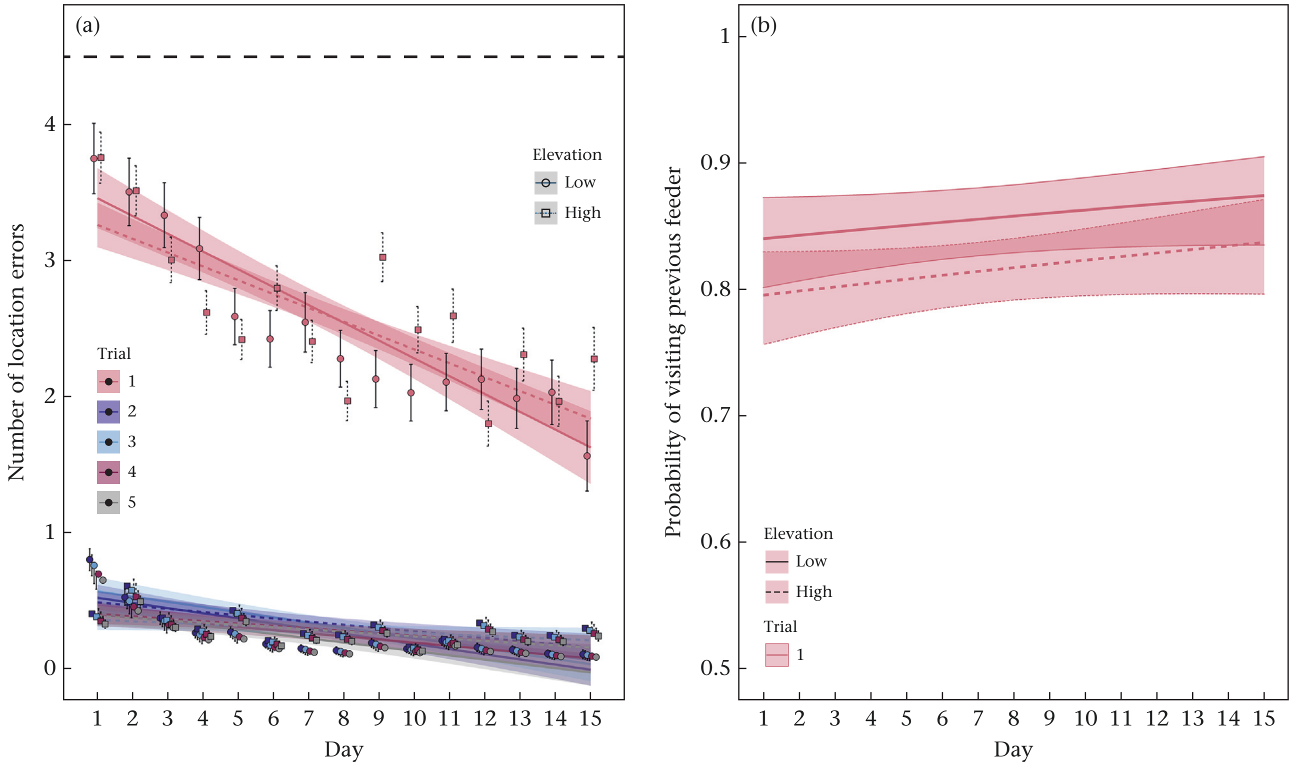

Figure 6. For Serial R2, (a) the number of location errors per trial for trials 1–5 per day and (b) the probability of visiting the previous feeder (i.e. making a reversal error) in trial 1 per day. Confidence intervals (shaded) generated from ‘ggplot lm’ method. Estimated marginal means (points) and standard errors (lines) were generated from reported GLMMs. Heavy dashed line indicates expected random value. In (b), predicted probabilities were simulated using GLMMs and ‘ggeffects’.

subsequent trials, making an average ± SE of 0.22 ± 0.73 location errors across trials 2–5 on the last day of Serial R2 (Fig. 6a). In contrast to Serial R1, performance improved significantly across days and differed significantly between elevations (Table 5). There was a significant interaction between elevation and day: low-elevation birds performed significantly better than high-elevation birds (Table 5). Similarly to Serial R1, birds that completed more previous reversal trials performed significantly worse during trials 2–5, while spatial learning and memory ability and single reversal ability were not significant predictors (Table 6).

Serial R2, last five trials

The sum of location errors made in the last five trials per day was significantly better than would be expected by chance (one-sample Wilcoxon signed-rank test: $V = 0$, $N = 235$, $P < 0.001$), but in contrast to Serial R1, it decreased significantly across days and did not vary significantly between elevations (Table 3, Appendix, Fig. A5). Birds made an average ± SE of 0.48 ± 0.88 total location errors in the last five trials per day across the Serial R2, which is consistent with better than 90% learning criterion (<0.5 errors/5 trials). Consistent with Serial R1, better performance in the last five trials of Serial R2 was associated with more trials completed in the same day (Table 3).

Serial R2: Learning to Predict

Unlike Serial R1, the probability of making a reversal error in trial 1 did not change significantly with days across Serial R2 ($N = 235$; Table 4, Fig. 6b). Birds at high elevation were significantly less likely to make a reversal error than birds at low elevation ($N = 235$; Table 4). Whereas birds with better spatial learning and memory ability were more likely to make a reversal error in trial 1 ($N = 180$; Table 4, Appendix, Fig. A6), there was no significant effect of single reversal ability ($N = 164$; Table 4). Consistent with Serial R1, birds that completed more previous reversal trials were also more likely to make a reversal error in trial 1 each day of Serial R2.

Exploring Proactive Interference in Serial R1 and R2

Neutral errors before reversal errors in trial 1

In Serial R1, during trials in which birds visited the feeder that was rewarding before the switch (i.e. made a reversal error), the number of neutral errors birds made before making a reversal error in trial 1 was very low (average ± SE = 0.04 ± 0.24 errors; $N = 97$) and did not vary across days, between elevations, with spatial learning and memory ability or with single reversal ability (Appendix, Table A4, Fig. A4). In contrast, during Serial R2, although the number of neutral errors before a reversal error in trial 1 was also very low (average ± SE = 0.07 ± 0.31; $N = 235$), birds made fewer such errors across days. Also, birds at low elevation and birds with better spatial learning and memory ability made significantly fewer neutral errors before making a reversal error compared to birds at high elevation and birds with worse spatial learning and memory (Appendix, Table A4, Fig. A4). Consistent with Serial R1, there was no significant effect of single reversal ability (Appendix, Table A4). Trials in which birds did not make a reversal error were excluded from both analyses.

Comparing performance between Serial R1 and R2

Birds that participated in both serial reversals ($N = 86$) in general made significantly more trial 1 location errors in Serial R2 both compared to birds that did not participate in Serial R1 at all ($N = 33$, Fig. 7b, Table 6) and compared to their performance in trial 1 of Serial R1 (Fig. 7a, Table 6).

Sum of reversal errors in the first 20 trials

In Serial R1, the sum of reversal errors in the first 20 trials per day did not vary between elevations, across days or with the number of previous reversal trials (Fig. 8, Table 7; $N = 97$). In contrast, in Serial R2, the sum of reversal errors in the first 20 trials decreased significantly with days across the task and was greater for birds that completed more previous reversal trials (Fig. 9, Table 7; $N = 86$). In this analysis, we included two types of reversal errors made in the same Serial R2 task: reversal errors made to Serial R1 feeder locations and reversal errors made to the reversal feeder location in Serial R2. But there was not a difference in error type on the sum of reversal errors (Table 7).

In Serial R2, the sum of reversal errors in the first 20 trials was significantly greater at high elevation and decreased significantly across days for both reversal errors made to Serial R2 locations and to Serial R1 locations but only varied significantly with previous reversal trials for Serial R2 reversal errors (Table 7).

For reversal errors made in Serial R2 to locations used in Serial R1, there was a significant interaction between elevation and day. Post hoc GLMMs suggested that these reversal errors decreased faster at low elevation and were not affected by the number of previous reversal trials (Fig. 9; Poisson GLMM: days: estimate ± SE = −0.48 ± 0.04, $\chi^2_1 = 138.95$, $P < 0.001$; number of previous reversal trials: estimate ± SE = 0.08 ± 0.09, $\chi^2_1 = 0.82$, $P = 0.36$; R^2c = 0.45, R^2m = 0.24; $N = 40$), compared to high elevation, where the number of previous reversal trials was

Table 6
GLMM model results comparing location errors in trial 1 of Serial R2 based on participation in Serial R1 and R2

Data set	Comparing Serial R1 and R2				
	Effects	Estimate	SE	χ^2_1	P
Full participation in Serial R1 and R2 or only full participation in Serial R2 ($N = 119$)	Day	−0.23	0.03	80.31	**<0.001**
	Elevation	<0.001	0.04	0.001	0.99
	SR1 Participation (full)	0.13	0.05	7.69	**0.006**
	Previous reversal trials	0.11	0.04	9.69	**0.002**
	Day∗elevation	0.07	0.03	4.07	**0.04**
	Bird ID (random intercept)	—	—	—	**Var. = 0.02**
				R^2c = 0.15, R^2m = 0.10	
Full participation in both Serial R1 and R2 ($N = 86$)	Day	−0.13	0.02	34.41	**<0.001**
	Elevation	−0.37	0.07	28.14	**<0.001**
	Task (Serial R1 or R2)	0.14	0.05	9.39	**0.002**
	Previous reversal trials	0.09	0.04	6.93	**0.008**
	Day∗elevation	−0.08	0.03	4.92	**0.03**
	Task∗elevation	0.38	0.1	26.80	**<0.001**
	Bird ID (random intercept)	—	—	—	**Var. = 0.02**
				R^2c = 0.16, R^2m = 0.11	

Conditional and marginal R^2 (R^2c and R^2m) were calculated according to Nakagawa and Schielzeth (2013). Elevation estimates are relative to high elevation. Bold text indicates P values less than the significance level ($\alpha = 0.05$).

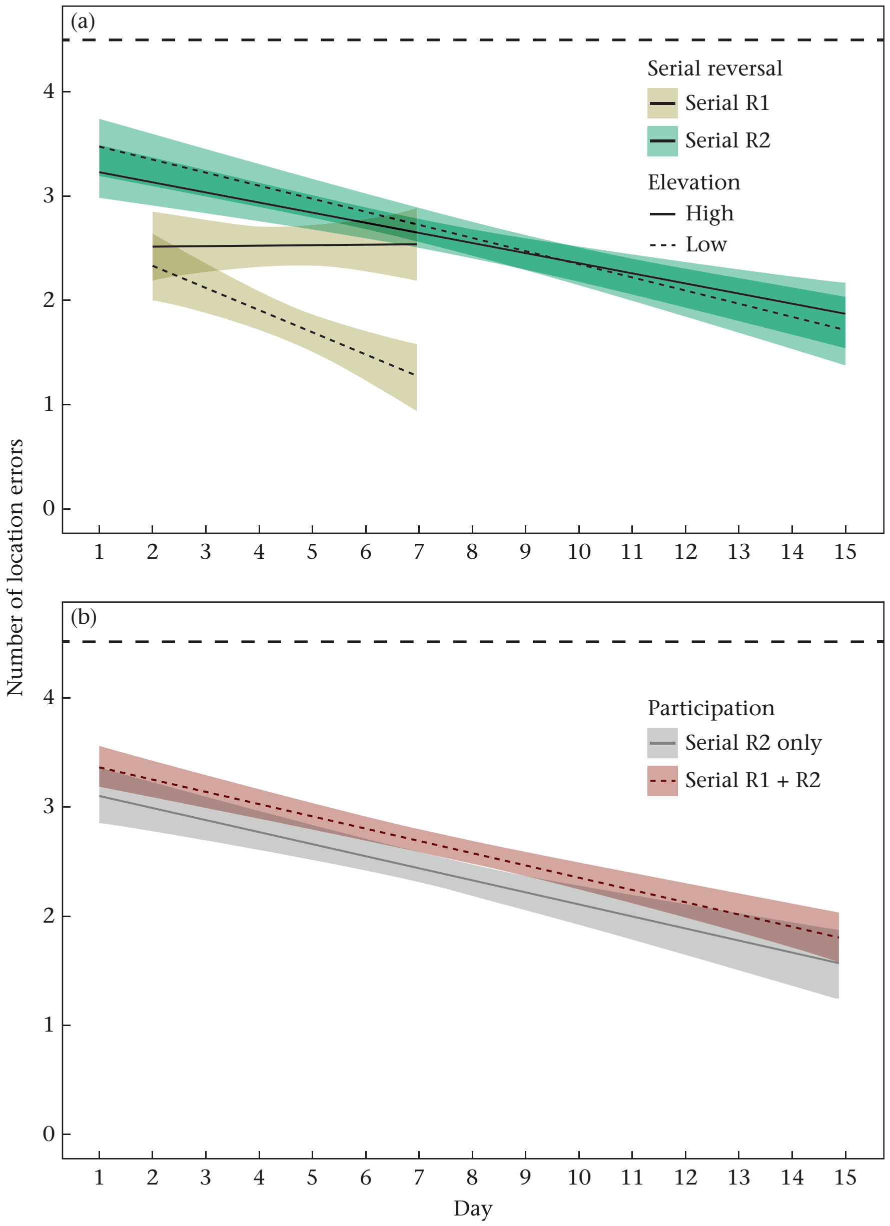

Figure 7. Performance (location errors) in trial 1 each day, comparing Serial R1 and R2, for (a) birds that participated in both Serial R1 and R2 ($N = 86$) by elevation and serial reversal task and (b) birds that participated in both Serial R1 and R2 ($N = 86$) compared to birds that only participated in Serial R2 ($N = 33$) by elevation. In (b), only Serial R2 performance is depicted. Heavy dashed line indicates expected random values.

positively correlated with reversal errors to Serial R1 locations (Poisson GLMM: days: estimate ± SE = -0.30 ± 0.04, $\chi^2_1 = 67.61$, $P < 0.001$; previous reversal trials: estimate ± SE = 0.19 ± 0.09, $\chi^2_1 = 3.86$, $P = 0.049$; R^2c = 0.40, R^2m = 0.11; $N = 46$).

For reversal errors made in Serial R2 to Serial R2 reversal feeders, there was also a significant interaction between elevation and day. Post hoc GLMMs indicated that at low elevation, these reversal errors decreased across days and were positively correlated with previous reversal trials (Poisson GLMM: days: estimate ± SE = -0.14 ± 0.04, $\chi^2_1 = 14.07$, $P < 0.001$; previous reversal trials: estimate ± SE = 0.25 ± 0.07, $\chi^2_1 = 11.20$, $P < 0.001$; R^2c = 0.06, R^2m = 0.05; $N = 40$), but there was no significant decrease in the number of these errors at high elevation (Poisson GLMM: days: estimate ± SE = -0.02 ± 0.03, $\chi^2_1 = 0.28$, $P = 0.60$; number of previous reversal trials: estimate ± SE = 0.02 ± 0.06, $\chi^2_1 = 0.08$, $P = 0.77$; $N = 46$). However, the high-elevation post hoc model had a singular fit, likely due to low variance among birds used as the random intercept ($\sigma^2 < 0.001$).

Supplemental Analyses of Neutral Errors

We also include results of analyses using only neutral errors instead of total location errors to evaluate performance in trial 1 and trials 2–5 for both Serial R1 and R2 (see Appendix, Table A6). These results are consistent with our main conclusion and do not suggest that neutral errors were correlated with spatial learning and memory performance.

DISCUSSION

The two main aims of the study were to test whether chickadees in the wild could (1) learn an eight-position spatial serial

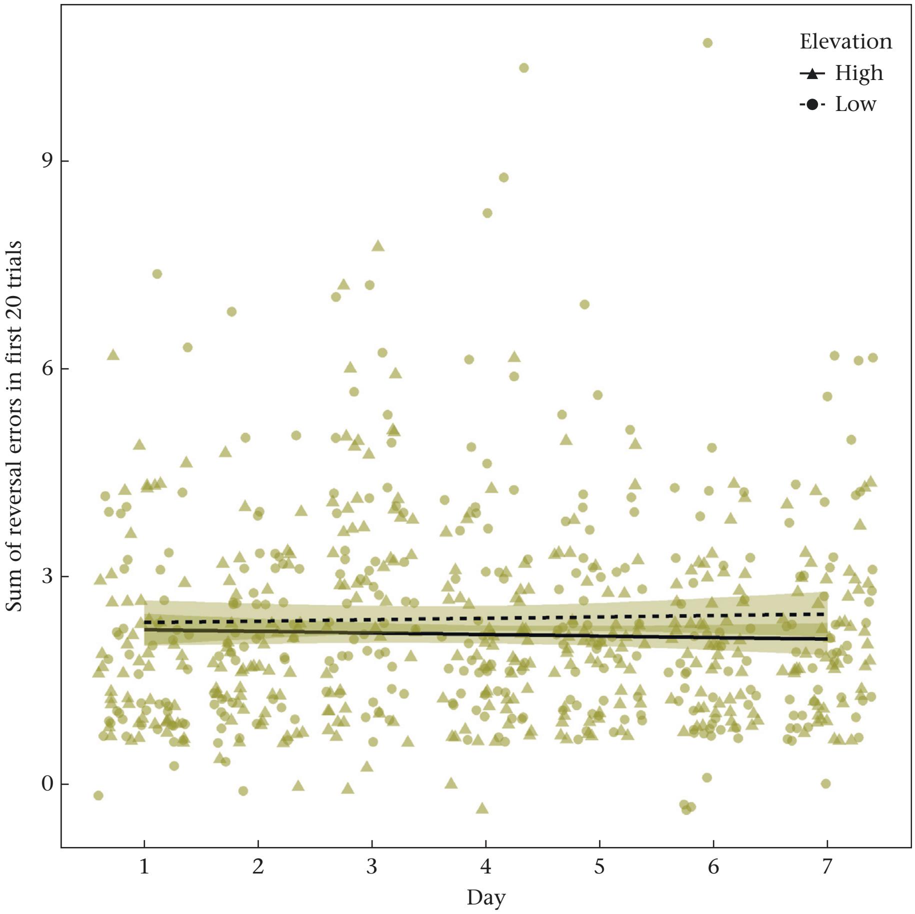

Figure 8. Sum of reversal errors in the first 20 trials per day of Serial R1 by day and elevation. All reversal errors were made during Serial R1. Includes only birds that participated in Serial R1 and the spatial learning and memory task and single reversal learning task ($N = 97$).

reversal learning task with daily location reversals by reaching performance consistent with learning the reversal rule and (2) learn to predict the next daily food location based on experience during the serial reversal task. Performance of birds at both elevations was consistent with learning the reversal rule and low-elevation birds consistently outperformed high-elevation birds. Chickadees learned to reverse associations in the two serial reversal tasks, making fewer and fewer location errors in trial 1 each day and making almost no location errors (including no neutral errors) after trial 2 each day. After just seven reversals in Serial R1 and 15 reversals in Serial R2, average performance in trial 1 after the reversal was close to a mean of 1.5 location errors at low elevation and a mean of 2.25 location errors at high elevation. Moreover, the proportion of birds showing trial 1 performance consistent with learning the reversal rule (a single reversal error and no neutral errors, associated with 'win-stay; lose-shift') increased at both elevations by the end of both serial reversal tasks. Again, this proportion was greater at low elevation than at high elevation. At the same time, birds clearly learned the rewarding feeder location by the end of every day in both tasks, as indicated by fewer than 0.5 mean location errors per trial during the last five trials of the day. This is significantly better than expected by chance and corresponds to better than a 90% learning criterion. Together, these results show that many birds learned to first visit the reversal feeder and then immediately sought out the correct rewarding feeder in the first trial after each reversal, which is consistent with learning the serial reversal rule (Chittka, 1998; Strang & Sherry, 2014).

Improvement in serial reversal performance was mainly due to a reduction in neutral errors during the first trial after each daily switch. In both tasks, birds reduced the total number of location errors in trial 1 across days. Considering that birds could make a maximum of one reversal error per trial, this suggests that birds made fewer neutral errors as they learned the task. In other words, they sampled fewer nonrewarding feeders while searching for the correct rewarding feeder each day across the task. Neutral errors during reversal learning show that birds did not fully retain memories of the previously rewarding feeder following a full day of reinforcement at a different feeder. Moreover, when birds made a reversal error in the first trial each day, they rarely made any neutral errors beforehand. This means that birds typically remembered the most recently visited feeder, and that neutral errors estimated how well they learned to remember the least recent (correct) feeder location. As birds made fewer neutral errors, they learned to remember both feeder locations and flexibly shift between those memories without inhibiting them. This is consistent with the increase in the proportion of birds showing 'one-trial, one-reversal-error' performance by the end of both serial reversal tasks. These results also appear consistent with cognitive flexibility, which is associated with executive function (e.g. Lai et al., 1995) and may allow birds to quickly switch learning contingencies while retaining previous memories.

An alternative explanation for the reduction in neutral errors is that neutral errors could reflect spatial learning and memory ability rather than reversal learning ability. However, if that were the case, then we would expect that birds with better spatial learning

Table 7
Sum of reversal errors in the first 20 trials

Response variable	Effects	Estimate	SE	χ^2_1	P
Serial R1: Sum of reversal errors in the first 20 trials per day ($N = 97$)	Elevation	0.13	0.07	3.44	0.06
	Day	−0.03	0.04	0.62	0.43
	Previous reversal trials	−0.02	0.06	0.09	0.77
	Elevation∗day	0.01	0.06	0.01	0.91
	Bird ID (random intercept)	—	—	—	$\sigma^2 = 0.046$
				$R^2\text{c} = 0.125$, $R^2\text{m} = 0.013$	
				AIC = 1987.9	
Serial R2: Sum of reversal errors in first 20 trials per day ($N = 86$)	Elevation	0.24	0.047	25.91	**<0.001**
	Days	−0.12	0.036	10.69	**0.001**
	Previous reversal trials	0.11	0.046	5.72	**0.02**
	Elevation∗day	0.1	0.046	4.94	**0.03**
	Bird ID (random intercept)	—	—	—	$\sigma^2 = 0.003$
				$R^2\text{c} = 0.05$, $R^2\text{m} = 0.05$	
				AIC = 3294.7	
Serial R2: Sum of reversal errors to Serial R1 locations in first 20 trials per day ($N = 86$)	Elevation	−0.01	0.12	0.003	0.95
	Day	−0.49	0.04	151.48	**<0.001**
	Previous reversal trials	0.12	0.06	3.79	0.05
	Elevation∗day	0.18	0.06	10.92	**<0.001**
	Bird ID (random intercept)	—	—	—	$\sigma^2 = 0.23$
				$R^2\text{c} = 0.42$, $R^2\text{m} = 0.17$	
				AIC = 3563.6	

Conditional and marginal R^2 (R^2c and R^2m) were calculated according to Nakagawa and Schielzeth (2013). Unstandardized regression coefficients used as estimates. Elevation estimates are relative to low elevation for Serial R1 and high elevation for Serial R2. For Serial R2, only including birds that fully participated in both Serial R1 and R2 ($N = 86$). Bold text indicates P values less than the significance level ($\alpha = 0.05$).

abilities would make fewer neutral errors throughout the task. To address this alternative, we measured spatial learning and memory ability before the serial reversal task. Our previous work shows that variation in this ability is associated with significant differences in overwinter survival (Sonnenberg et al., 2019) and has a genetic basis (Branch et al., 2022). Yet variation in spatial learning and

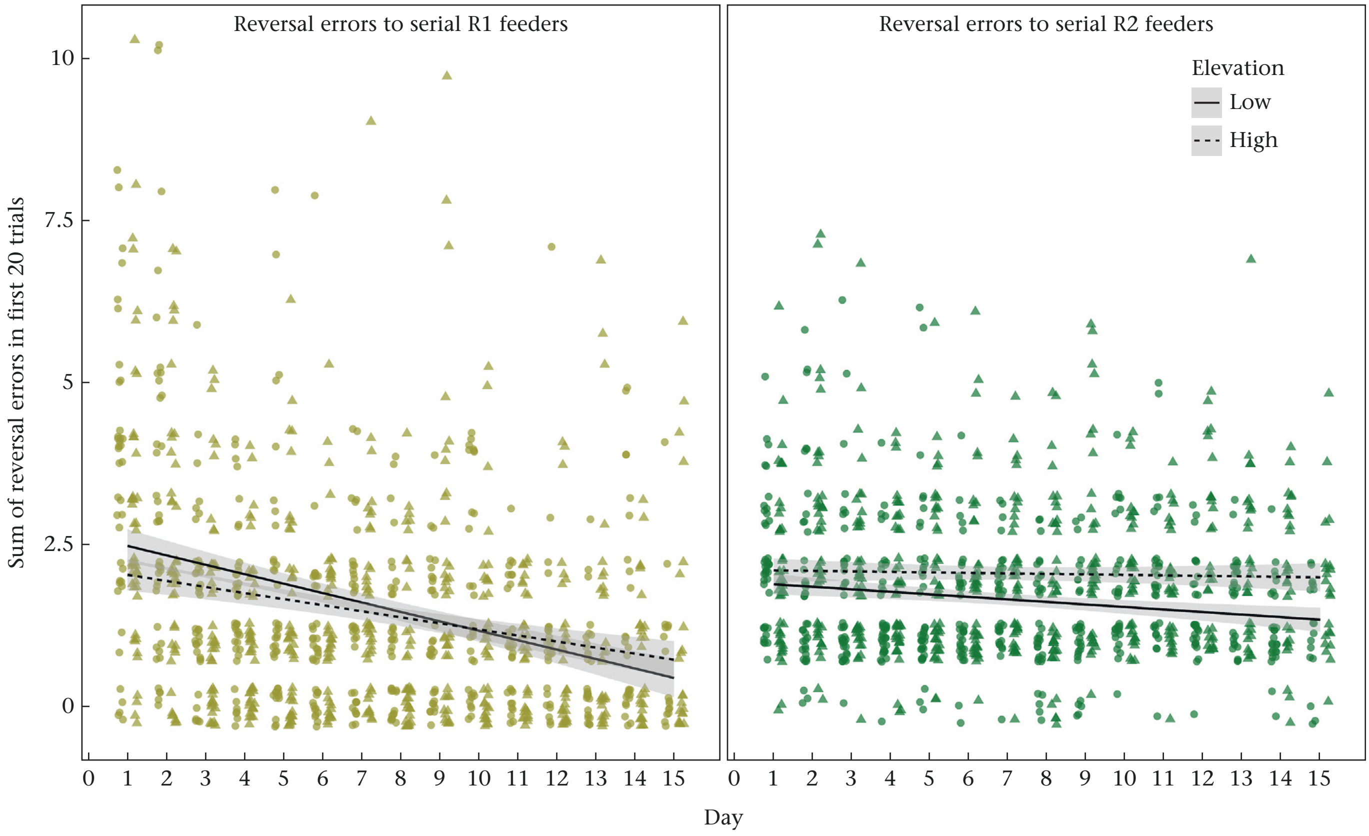

Figure 9. Sum of reversal errors in the first 20 trials per day of Serial R2 by day elevation (high: triangles; low: circles) and reversal error type. All reversal errors were made during trials in Serial R2, separated by type: reversal errors to Serial R1 target feeders (yellow) and reversal errors made to the Serial R2 target feeders (green). Includes only birds that participated in Serial R1 and Serial R2 and the spatial learning and memory task and the single reversal learning task ($N = 86$).

memory ability was not significantly associated with differences in neutral errors during the serial reversal task, whether tested as total location errors or separately. Additionally, we found elevational differences in serial reversal performance that were opposite from previous results for spatial learning and memory; specifically, our present results show that birds from low elevation outperformed high-elevation birds for serial reversal, yet we have previously shown that high-elevation birds have better spatial learning and memory abilities than birds at low elevation (Croston et al., 2016, 2017; Freas et al., 2012). We also find the alternative explanation unlikely because learning within a reversal paradigm assumes interference from previous memories, and thus makes serial reversal performance distinct from basic learning and memory ability. Flexibility is measured by how fast animals can learn changing associations, and in the case of serial reversal, how fast animals can learn repeatedly changing learning contingencies. To measure the rate of serial reversal learning is to measure the reduction in all possible errors across serial reversals.

We also found that learning the serial reversal in Serial R1 was not necessarily determined by persistence to the reversal feeder. Although the number of reversal errors that birds made during the first 20 trials per day did not change at either elevation across Serial R1, birds at low elevation still significantly improved performance in trial 1 throughout the task. This shows that such learning can occur independently of changes in reversal errors. Considering that trial 1 was the first trial after the rewarding feeder location switches, at this point birds did not have experience with the new day's rewarding feeder. As such, improved performance in trial 1 suggests that birds were learning the reversal rule, not relearning the associations each day. During Serial R2, overall improvement in serial reversal performance was associated with both a reduction in the number of neutral errors during trial 1 and a reduction in the sum of reversal errors over the first 20 trials, with high-elevation birds showing worse performance in both compared to low-elevation birds. It is not clear why we observed such differences between Serial R1 and Serial R2, but it is likely related to differences in timing to learn the reversing locations (e.g. 5 days versus 1 day) before each task.

We found mixed support for chickadees' ability to learn to predict the next daily food location based on experience during the serial reversal task by the end of Serial R1 at low elevation (ca. 21% of birds demonstrated an ability to predict which feeder would provide food the next day). As Serial R1 progressed, low-elevation birds overall made fewer location errors in trial 1 and were less likely to make a reversal error, instead making no or few neutral errors in trial 1. Together, these results suggest that these birds both learned the daily feeder switching pattern and were able to predict which feeder would provide food the next day based on their memory of the previously rewarded feeders in the serial reversal. Being able to predict future locations supports the hypothesis that chickadees learned the reversal rule and then used that rule to predict the daily pattern of switching. Considering that these birds were able to learn to predict the reversal pattern in just seven reversals, it is likely that given more time, more birds might be expected to show this behaviour. In contrast, there was no evidence that high-elevation birds learned to predict the reversal pattern despite many showing 'perfect' performance consistent with learning the reversal rule and only making a single error during trial 1 each day. These results suggest that high-elevation chickadees are overall less cognitively flexible than low-elevation birds.

There was also no evidence that birds were able to learn to predict the correct rewarding feeder at either elevation during Serial R2, potentially due to the increased difficulty of the task. Serial R2 used two new rewarding locations, requiring birds that had already participated in Serial R1 to learn two additional locations. Additionally, in the first task, birds had 5 days to learn the association with each feeder location used in Serial R1, whereas in the second task, birds only had 1 day to learn each location before starting Serial R2. Considering that proactive interference may be stronger with more recent memories (Storm & Bjork, 2016), we think that providing 5 days to learn each association may have increased the retention interval (i.e. time) between learning the first feeder location and beginning to reverse. In this case, the 1-day retention interval in Serial R2 would be shorter than the 5-day retention interval for Serial R2, potentially leading to stronger proactive interference when the reversals began and making Serial R2 more difficult. Accordingly, across days in Serial R2, birds also improved performance in the last five trials per day and made fewer reversal errors in the first 20 trials each day. These improvements contrast with Serial R1, during which birds at low elevation showed no improvement in the number of reversal errors across days while greatly reducing the number of neutral errors.

To our knowledge, this is the first study to explore a multiple position serial reversal task using spatial reversals associated with changing days under natural conditions. Birds in this study learned both serial reversal tasks quickly in just seven and 15 daily reversals in each task. Birds also reduced location errors dramatically within each reversal: in just the first two to five trials of the first reversal (day 1), birds made less than 0.25 mean location errors per trial in Serial R1 and less than 0.75 mean location errors per trial in Serial R2. Such performance is in stark contrast to many laboratory-based studies showing that other species require many more trials to learn a serial reversal task (Bond et al., 2007; Bublitz et al., 2017; Cauchoix et al., 2017; Hermer et al., 2018, 2021; Liu et al., 2016; Mackintosh & Cauty, 1971; Madden et al., 2018; van Horik et al., 2019) and rarely reach the levels of performance observed in wild chickadees. However, these differences may be attributed to differences in study design: in these previous studies, each reversal switched after the individual reached a learning criterion and individuals only had two choices that were closely situated. These differences in design mean that individuals in these studies completed fewer trials per reversal compared to our study. Plus, the costs of sampling all choices would likely be lower in the binary choice tasks compared to our eight-position task. Serial reversal performance may also differ in difficulty across spatial and nonspatial contexts, resulting in further differences between our results and the literature. However, we find these explanations unlikely. Completing more trials per reversal, as in our study, could result in stronger associations and higher levels of proactive interference, making our task more difficult than those in the literature. In food-caching chickadees, spatial associations may also be more difficult to learn than cue-based associations (Pravosudov & Clayton, 2001, 2002). Yet wild chickadees in their natural environment showed faster learning compared to previous laboratory studies, suggesting the exceptional performance of chickadees in our study was likely not due to the study design.

Considering that by the end of both reversal tasks, almost half the birds showed 'one-trial, one-error' reversal performance, the performance of chickadees on these serial reversal tasks is consistent with learning the reversal rule (e.g. 'win-stay, lose-shift'). This rule is based on flexible memories, as an individual must remember both locations in order to first visit the most recently rewarding location and then find the currently rewarding location without making any other errors. The fact that some low-elevation birds learned to predict the future locations also suggests that the preceding ‘one-trial, one-error’ performance involves learning the rule. Learning to predict requires birds to first learn to keep multiple memories of switching locations without inhibiting them after each reversal and then learn the switching rule that can be used to make predictions.

Elevational Differences and Proactive Interference

We predicted that birds from a more predictable and milder winter environment (e.g. low elevation) would be better at learning the reversal rule and predicting the next reversal feeder than birds from a less predictable, harsher winter environment (e.g. high elevation). This prediction was based on previously documented differences in spatial learning and memory ability, memory retention (Freas et al., 2012) and memory load associated with differences in food caching (Freas et al., 2012). We found support for our prediction: (1) birds at low elevation consistently outperformed birds at high elevation during trial 1 of both serial reversal tasks, (2) a greater proportion of birds learned the reversal task faster at low elevation, (3) only birds at low elevation showed an ability to learn to predict the next rewarding feeder during Serial R1 and (4) birds at low elevation made fewer reversal errors during the first 20 trials during Serial R2. We have previously shown that birds at low elevation are more cognitively flexible on a single spatial reversal task compared to birds at high elevation during some years (Croston et al., 2017; Tello-Ramos et al., 2018). These data suggest that cognitive flexibility is associated with learning to predict changing locations across days. However, we did not detect differences between elevations in single reversal ability or learning and memory ability during this study. This may suggest that the differences in learning to reverse and predict between elevations could be due to other memory related-cognitive differences. This explanation is supported by previous work showing that birds at high elevation in our system are under strong natural selection for spatial learning and memory ability but not for spatial reversal learning ability and typically show better learning and memory performance than birds at low elevation (Croston et al., 2016, 2017; Freas et al., 2012; Sonnenberg et al., 2019). Alternatively, these elevational differences could be driven by differences in caching and the associated memory load (Tello-Ramos et al., 2019). Birds at high elevation typically cache more than birds at low elevation (Freas et al., 2012); therefore, these birds may have had higher memory load and longer memory retention (Freas et al., 2012), which could interfere more with learning a serial reversal task (Tello-Ramos et al., 2019). In addition, Tello-Ramos et al. (2019) suggested that food-caching species do not need to retain memories associated with retrieved caches; instead, they likely simply forget these locations, eliminating the need to keep updating their memory of retrieved caches. Such forgetting seems to be associated with hippocampal neurogenesis rates, which are higher in birds at high elevation (Freas et al., 2012; Tello-Ramos et al., 2019). Overall, elevational differences in performance appear to be consistent with differences in proactive interference associated with previous memories negatively affecting repeatedly learning and recalling new information.

Assessing Support for Proactive Interference

We think several key results suggest that proactive interference may affect performance and explain differences in performance between serial reversal tasks and elevations. First, birds that completed more trials in the previous reversal during Serial R2 (1) made more location errors, (2) were more likely to make a reversal error in trial 1 each day and (3) made more reversal errors in the first 20 trials per day. This suggests that more experience learning the reversal association was correlated with both worse performance while learning the new association in the subsequent reversal and more persistence in visiting the reversal feeder. This is exactly in line with predictions of proactive interference (Tello-Ramos et al., 2019). Furthermore, our results are consistent with higher levels of proactive interference at high elevation than at low elevation: in Serial R1 we only see the association between the number of previous reversal trials and making a reversal error in trial 1 at high elevation. In contrast, some birds at low elevation, but none at high elevation, were also able to predict the next reversal location and may have demonstrated higher cognitive flexibility.

Furthermore, performance in Serial R2 was overall worse than in Serial R1. One explanation for this may be the difference in time to learn the first reversal: 1 day in Serial R2 compared to 5 days in Serial R1. More time to learn the single reversal may have allowed individuals in Serial R1 to make an association with each rewarding location but also have more time to unlearn or inhibit the recall of the reversal feeder location. Both stronger memories and more time to overcome the initial interference may have helped birds reform the associations between cues and memories in order to learn the reversal rule.

Another alternative explanation for these differences between reversal tasks could be that birds responded to the snowstorm that interrupted Serial R1 and R2. If this storm increased birds' perception of environmental harshness or unpredictability, birds may have relied more on previous memories to find the rewarding feeder. Or, under this paradigm, it may be more adaptive under harsh or changing conditions to sample previously rewarding feeders, leading to worse performance during Serial R2. However, our previous work shows that chickadees will reduce, rather than increase, sampling of food sources that have stopped rewarding (Benedict et al., 2021). Furthermore, we find the snowstorm explanation less likely because we would not expect cognitive traits to vary widely with immediate environmental conditions. We also would have then expected to see a more dramatic decrease in performance at high elevation, where conditions were more affected by the snowstorm.

Birds without any Serial R1 experience performed better in trial 1 per day of Serial R2 compared to birds that participated in both serial reversal tasks, which is also consistent with a detrimental effect of proactive interference on learning the second serial reversal task. These differences also appear in line with proactive interference and memory load limitations (Hermer et al., 2021; Tello-Ramos et al., 2019), as birds that learned both tasks needed to learn more feeder locations than birds that only learned one task. We have previously shown evidence of proactive interference in chickadees at the population level using a single reversal task (Croston et al., 2017; Tello-Ramos et al., 2018), and a similar effect of proactive interference on reversal performance was shown in captive great tits, *Parus major*, on an individual level (Hermer et al., 2021).

In further support of the proactive interference hypothesis, our data on the sum of reversal errors in the first 20 trials per day provides an estimate of how birds persisted in visiting the reversal feeder when learning each reversal. For birds that participated in both serial reversal tasks, we measured Serial R2 reversal errors during the first 20 trials to both the reversal location in Serial R2 and to the two rewarding locations used in Serial R1. We found that low-elevation birds in Serial R2 made fewer reversal errors of both types across days, but high-elevation birds only reduced the number of reversal errors associated with Serial R1 and continued to make reversal errors to the reversal feeder in Serial R2. Overall, birds at high elevation made significantly more reversal errors of both types compared to birds at low elevation, which again is consistent with our hypothesis of elevation-related differences in proactive interference (Croston et al., 2017; Tello-Ramos et al., 2018, 2019).

A possible alternative to the proactive interference hypothesis is that birds made more reversal errors because persistence to visit previously rewarding feeders could be advantageous. Such persistence could be beneficial in environments with frequently or unpredictably replenishing resources, leading individuals to forage from familiar locations instead of using energy to explore new resources. A key aspect of this 'adaptive persistence' hypothesis is the idea that making reversal errors reflects an adaptive behaviour rather than a by-product of proactive interference. However, within this framework, we would expect chickadees to persist at unrewarding feeders even after those feeders stop providing food, particularly at high elevation where conditions are harsher. Yet, a previous study found the opposite: chickadees at both elevations significantly reduced visits to the previously rewarding feeders after the feeders stopped providing food, with no differences between elevations (Benedict et al., 2021). As such, further evidence is necessary to explore persistence in other contexts and study systems.

Overall, our results show that low-elevation birds are more cognitively flexible, as they were faster and better at learning the serial reversal task and at learning to predict the switching location based on a daily changing pattern, regardless of potential differences in proactive interference levels between elevations. We also show that at least some differences in serial reversal performance between elevations are consistent with differences in proactive interference levels. The only unexplained and contradictory result was that high-elevation birds were less likely to make a reversal error during trial 1 of the Serial R2 compared to birds from low elevation. But this difference was reversed in trials 2–5, which, combined with all other evidence presented above, strongly suggests that the overall levels of proactive interference were higher at high elevation.

Age Not Correlated with Serial Reversal Performance

We did not find support for our prediction that adults would be more cognitively flexible compared to juveniles and would perform better on the serial reversal task. There were no significant associations between age and serial reversal performance or learning to predict. We also did not find a difference between age groups in single reversal ability (assessed before Serial R1); however, adults performed significantly better than juveniles while learning the locations for Serial R2. While this latter result is consistent with our previous results, suggesting that juveniles were slower to learn a single reversal than adults (Tello-Ramos et al., 2018), we are unsure why we did not see this effect on the single reversal task before Serial R1. On the other hand, chickadees of all ages (up to 8 years old) showed similar performance on the serial reversal, which is consistent with our previous work indicating that chickadees do not show age-related cognitive senescence up to 8 years of age (Heinen, Pitera, Sonnenberg, Benedict, Branch et al., 2021).

No Evidence of Anticipatory Errors

We found that birds made significantly fewer location errors than expected by chance during the last five trials per day of both Serial R1 and R2, suggesting that birds learned the feeder location by the end of their participation each day. This was important to confirm because we did not use a learning criterion to cue each reversal. But regardless, chickadees made, on average, less than 0.5 errors in the last five trials, which is consistent with better than 90% success. We also found that the number of location errors either did not change (Serial R1) or decreased (Serial R2) across days, providing evidence against anticipatory errors. As such, we do not believe birds associated the reversal switch with counting or timed intervals.

Learning to Predict and Implications for Mental Time Travel

The 'stuck-in-time' hypothesis argues that nonhuman animals cannot remember events associated with a specific point in time (i.e. episodic memory) or anticipate events far out in the future (Roberts, 2002; Roberts & Feeney, 2009). Yet, in our study, some birds at low elevation demonstrated an ability to predict a spatial location based on their previous experience within just seven reversals in 7 days. Although many birds continued to visit the previous feeder first in trial 1, apparently relying on memory instead of learning to predict, approximately 21% of birds at low elevation in Serial R1 went directly to the correct feeder during their first trial without making a reversal error. While these birds first learned the reversal rule to switch feeders, our results suggest they may have also associated this rule with time across days in order to predict the next feeder location. Such an abstract association may involve a higher-level cognitive process to integrate memories of where they found food the previous day, where they found food 2 days prior and that the food location alternated between these two locations each day. Each day, a bird should remember which feeder was rewarded the previous day to make a correct prediction about the location of the currently rewarding feeder. Although our study design did not provide evidence for a specific mechanism, our data show that these birds had to associate previous experiences with several points in time to correctly predict where the feeder would be and decide which feeder to visit first the next day. Even though our study was conducted with wild birds in their natural environment without strict control groups, it suggests that chickadees may not be 'stuck in time'.

We think our results suggest that a food-caching bird may have used mental time travel (MTT) to associate the reversal rule with days. MTT is a cognitive process in which individuals use semantic memory (knowledge of facts) and episodic memory (remembering past experiences) to reconstruct memories of past events that occurred at a particular point in time (Roberts, 2002). MTT can be used to represent future events that have not occurred yet, allowing individuals to flexibly project their past experiences to simulate potential outcomes of future events (Suddendorf & Corbalis, 2007). While it is extremely difficult to study MTT in nonhuman animals, especially in the field, we think our study design allowed us to investigate several abilities that would be required for chickadees to be able to use MTT. Previous work has shown that black-capped chickadees, *Poecile atricapillus*, are capable of the episodic-like memory and can use semantic knowledge from previous experience to predict future needs and plan accordingly (Feeney et al., 2011). Our study builds on this work, suggesting that mountain chickadees can (1) associate two feeder locations with two different points in time and (2) can use that information to decide their future action outside of a 24 h period. Both these behaviours have been associated with mental time travel (Cheke & Clayton, 2010; Roberts & Feeney, 2009; Suddendorf & Corbalis, 2007).

There are potential alternative explanations for our results, such as using simple rules to track time. For example, birds may choose to visit the feeder that is least recent in their memory (e.g. Roberts et al., 2008), even though it is unclear what cue they might use to switch locations each day. Or birds may learn that the feeder location switches after roughly a day, but without associating this change with their specific past experience of foraging at the feeder (Cheke & Clayton, 2010). Another possibility is that the birds may be using a simpler rule to predict the next location each day (e.g. counting time between their last visit per day and their first visit the next day), but we find this unlikely because there was variation in how frequently birds visited the feeders each day. Specifically, time intervals between trials ranged from seconds to hours and different birds visited the feeders at different times of day, so it is unlikely that all birds were associating the switching food reward with the same time cue (i.e. dawn). We also did not observe any anticipatory errors to the next rewarding feeder in the last five trials of each day, which would be likely if birds had associated the switching feeder with counting or an imprecise circadian rhythm (e.g. Rayburn-Reeves et al., 2011).

Conclusions

Overall, our data generally supported our main predictions that (1) chickadees are highly cognitively flexible and can learn a serial reversal task across days under natural conditions, reaching perfect ‘one-trial, one-error’ performance consistent with learning the reversal rule; (2) chickadees from low elevation with a milder and more predictable environment would learn the reversal task and reach the perfect ‘one-trial, one-error’ performance significantly faster compared to birds from a harsher and more unpredictable high elevation and (3) chickadees are capable of learning to predict a rewarding feeder after learning the reversal rule across multiple days, albeit only at low elevation. Our data are largely consistent with the effects of proactive interference on serial reversal learning, in which strong spatial memories and the strength of learned associations may negatively affect cognitive flexibility. Furthermore, the ability to learn to predict a spatial location based on previous experience may involve cognitive flexibility and may suggest that these birds are able to associate a rule with more abstract concepts, such as time.

Author Contributions

Lauren M. Benedict: Conceptualization, Methodology, Investigation, Formal analysis, Writing – Original draft preparation, Writing – Reviewing and editing, Visualization. **Virginia K. Heinen**: Methodology, Investigation, Data curation, Writing – Reviewing and editing. **Benjamin R. Sonnenberg**: Investigation, Writing – Reviewing and editing. **Eli S. Bridge:** Methodology, Writing – Reviewing and editing. **Vladimir V. Pravosudov:** Conceptualization, Methodology, Investigation, Writing – Original draft preparation, Writing – Reviewing and editing, Funding acquisition, Project administration, Supervision.

Data Availability

Data will be made available on request.

Declaration of Interest

None.

Acknowledgments

This work was supported by the U.S. National Science Foundation (NSF IOS1856181 and IOS2119824 to V.V.P.). L.M.B. and B.R.S. were supported by the NSF Graduate Research Fellowship Program. Thank you to the staff of Sagehen Experimental Forest and Sagehen Creek Field Station (University of California Berkeley) for assistance. We also thank Carrie L. Branch and Angela M. Pitera for feedback during the writing process. Constructive comments from two referees significantly improved the manuscript. We particularly appreciate thought-provoking and constructive comments from Dr Arnon Lotem, which forced us to think critically about our data and interpretation of our results.

References

Allen, M., Poggiali, D., Whitaker, K., Rhys Marshall, T., & Kievit, R. A. (2021). Raincloud plots: A multi-platform tool for robust data visualization [version 2; peer review: 2 approved]. *Wellcome Open Research, 4*, 63. https://doi.org/10.12688/wellcomeopenres.15191.2

Anderson, M. C., & Neely, J. H. (1996). Interference and inhibition in memory retrieval. In E. L. Bjork, & R. A. Bjork (Eds.), *Memory* (pp. 237–313). Academic Press. https://doi.org/10.1016/B978-012102570-0/50010-0.

Aoki, K., & Feldman, M. W. (2014). Evolution of learning strategies in temporally and spatially variable environments: A review of theory. *Theoretical Population Biology, 91*, 3–19.

Audet, J., & Lefebvre, L. (2017). What's flexible in behavioral flexibility? *Behavioral Ecology, 28*, 943–947. https://doi.org/10.1093/beheco/arx007

Badre, D., & Wagner, A. D. (2006). Computational and neurobiological mechanisms underlying cognitive flexibility. *Proceedings of the National Academy of Sciences of the United States of America, 103*(18), 7186–7191. https://doi.org/10.1073/pnas.0509550103

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Machler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*, 1–48. https://doi.org/10.18637/jss.v067.i01

Bebus, S. E., Small, T. W., Jones, B. C., Elderbrock, E. K., & Schoech, S. J. (2016). Associative learning is inversely related to reversal learning and varies with nestling corticosterone exposure. *Animal Behaviour, 111*, 251–260.

Benedict, L. M., Pitera, A. M., Branch, C. L., Sonnenberg, B. R., Heinen, V. K., Bridge, E. S., & Pravosudov, V. V. (2021). Information maintenance of food sources is associated with environment, spatial cognition and age in a food-caching bird. *Animal Behaviour, 182*, 153–172.

Biebach, H., Gordijn, M., & Krebs, J. R. (1989). Time-and-place learning by garden warblers, *Sylvia borin*. *Animal Behaviour, 37*, 353–360.

Bjork, R. A. (1989). Retrieval inhibition as an adaptive mechanism in human memory. In H. L. Roediger, III, & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 309–330). L. Erlbaum.

Bond, A. B., Kamil, A. C., & Balda, R. P. (2007). Serial reversal learning and the evolution of behavioral flexibility in three species of North American corvids (*Gymnorhinus cyanocephalus, Nucifraga columbiana, Aphelocoma californica*). *Journal of Comparative Psychology, 121*(4), 372–379. https://doi.org/10.1037/0735-7036.121.4.372

Branch, C. L., Semenov, G. A., Wagner, D. N., Sonnenberg, B. R., Pitera, A. M., Bridge, E. S., Taylor, S. A., & Pravosudov, V. V. (2022). The genetic basis of spatial cognitive variation in a food-caching bird. *Current Biology, 32*, 210–219. https://doi.org/10.1016/j.cub.2021.10.036

Bridge, E. S., & Bonter, D. N. (2011). A low-cost radio frequency identification device for ornithological research. *Journal of Field Ornithology, 82*(1), 52–59. https://doi.org/10.1111/j.1557-9263.2010.00307.x

Bublitz, A., Weinhold, S. R., Strobel, S., Dehnhardt, G., & Hanke, F. D. (2017). Reconsideration of serial visual reversal learning in octopus (*Octopus vulgaris*) from a methodological perspective. *Frontiers in Physiology, 8*. https://doi.org/10.3389/fphys.2017.00054. Article 54.

Cauchoix, M., Hermer, E., Chaine, A. S., & Morand-Ferron, J. (2017). Cognition in the field: Comparison of reversal learning performance in captive and wild passerines. *Scientific Reports, 7*(1). https://doi.org/10.1038/s41598-017-13179-5. Article 12945.

Cheke, L. G., & Clayton, N. S. (2010). Mental time travel in animals. *Wiley Interdisciplinary Review of Cognitive Science, 1*, 915–930. https://doi.org/10.1002/wcs.59

Chittka, L. (1998). Sensorimotor learning in bumblebees: Long-term retention and reversal training. *Journal of Experimental Biology, 201*, 515–524. https://doi.org/10.1242/jeb.201.4.515

Croston, R., Branch, C. L., Pitera, A. M., Kozlovsky, D. Y., Bridge, E. S., Parchman, T., & Pravosudov, V. V. (2017). Predictably harsh environment is associated with reduced cognitive flexibility in wild food-caching mountain chickadees. *Animal Behaviour, 123*, 139–149. https://doi.org/10.1016/j.anbehav.2016.10.004

Croston, R., Kozlovsky, D. Y., Branch, C. L., Parchman, T. L., Bridge, E. S., & Pravosudov, V. V. (2016). Individual variation in spatial memory performance in wild mountain chickadees from different elevations. *Animal Behaviour, 111*, 225–234. https://doi.org/10.1016/j.anbehav.2015.10.015

Devine, J. V., Burke, M. W., & Rohack, J. J. (1979). Stimulus similarity and order as factors in visual short-term memory in nonhuman primates. *Journal of Experimental Psychology: Animal Behavior Processes, 5*, 335–354. https://doi.org/10.1037/0097-7403.5.4.335

Dridi, S., & Lehmann, L. (2016). Environmental complexity favors the evolution of learning. *Behavioral Ecology, 27*, 842–850. https://doi.org/10.1093/beheco/arv184

Dunlap, A. S., & Stephens, D. W. (2016). Reliability, uncertainty, and costs in the evolution of animal learning. *Current Opinion in Behavioral Sciences, 12*, 73–79.

Feeney, M. C., Roberts, W. A., & Sherry, D. F. (2011). Black-capped chickadees (*Poecile atricapillus*) anticipate future outcomes of foraging choices. *Journal of Experimental Psychology: Animal Behavior Processes, 37*, 30–40. https://doi.org/10.1037/a0019908

Freas, C. A., LaDage, L. D., Roth, T. C., II, & Pravosudov, V. V. (2012). Elevation-related differences in memory and the hippocampus in mountain chickadees, *Poecile gambeli*. *Animal Behaviour, 84*(1), 121–127. https://doi.org/10.1016/j.anbehav.2012.04.018

Harlow, H. F. (1949). The formation of learning sets. *Psychological Review, 56*, 51–65.

Hartig, F. (2020). *DHARMa: Residual diagnostics for hierarchical (multi-level/mixed) regression models* (Version 0.2.7) https://CRAN.R-project.org/package=DHARMa.

Heinen, V. K., Benedict, L. M., Pitera, A. M., Sonnenberg, B. R., Bridge, E. S., & Pravosudov, V. V. (2021). Social dominance has limited effects on spatial cognition in a wild food-caching bird. *Proceedings of the Royal Society B: Biological Sciences, 288*. https://doi.org/10.1098/rspb.2021.1784. Article 20211784.

Heinen, V. K., Pitera, A. M., Sonnenberg, B. R., Benedict, L. M., Branch, C. L., Bridge, E. S., & Pravosudov, V. V. (2021). Specialized spatial cognition is associated with reduced cognitive senescence in a food-caching bird. *Proceedings of the Royal Society B: Biological Sciences, 288*. https://doi.org/10.1098/rspb.2020.3180. Article 20203180.

Heinen, V. K., Pitera, A. M., Sonnenberg, B. R., Benedict, L. M., Bridge, E. S., Farine, D. R., & Pravosudov, V. V. (2021). Food discovery is associated with different reliance on social learning and lower cognitive flexibility across environments in a food-caching bird. *Proceedings of the Royal Society B: Biological Sciences, 288*. https://doi.org/10.1098/rspb.2020.2843. Article 20202843.

Hermer, E., Cauchoix, M., Chaine, A. S., & Morand-Ferron, J. (2018). Elevation-related difference in serial reversal learning ability in a nonscatter hoarding passerine. *Behavioral Ecology, 29*(4), 840–847. https://doi.org/10.1093/beheco/ary067

Hermer, E., Murphy, B., Chaine, A. S., & Morand-Ferron, J. (2021). Great tits who remember more accurately have difficulty forgetting, but variation is not driven by environmental harshness. *Scientific Reports, 11*. Article 10083.

Izquierdo, A., Brigman, J. L., Radke, A. K., Rudebeck, P. H., & Holmes, A. (2017). The neural basis of reversal learning: An updated perspective. *Neuroscience, 325*, 12–26.

Jacoby, L. L., Hay, J. F., & Debner, J. A. (2001). Proactive interference, accessibility bias, and process dissociations: Valid subjective reports of memory. *Journal of Experimental Psychology, 27*, 686–700.

Kozlovsky, D. Y., Branch, C. L., Pitera, A. M., & Pravosudov, V. V. (2018). Fluctuations in annual climatic extremes are associated with reproductive variation in resident mountain chickadees. *Royal Society Open Science, 5*. https://doi.org/10.1098/rsos.171604. Article 171604.

Lai, Z. C., Moss, M. B., Killiany, R. J., Rosene, D. L., & Herndon, J. G. (1995). Executive system dysfunction in the aged monkey: Spatial and object reversal learning. *Neurobiology of Aging, 16*, 947–954.

Lenth, R. (2020). *emmeans: Estimated marginal means, aka least-squares means* (Version 1.4.7) https://CRAN.R-project.org/package=emmeans.

Lewis, J. L., & Kamil, A. C. (2006). Interference effects in the memory for serially presented locations in Clark's nutcrackers, *Nucifraga columbiana*. *Journal of Experimental Psychology: Animal Behavior Processes, 32*, 407–418. https://doi.org/10.1037/0097-7403.32.4.407

Lewis, J. L., Kamil, A. C., & Webbink, K. E. (2013). Changing room cues reduces the effects on proactive interference in Clark's nutcrackers, *Nucifraga columbiana*. *Journal of Experimental Psychology, 39*, 187–192. https://doi.org/10.1037/a0031979

Liu, Y., Day, L. B., Summers, K., & Burmeister, S. S. (2016). Learning to learn: Advanced behavioural flexibility in a poison frog. *Animal Behaviour, 111*, 167–172. https://doi.org/10.1016/j.anbehav.2015.10.018 0003-3472/

Ludecke, D. (2018). ggeffects: Tidy data frames of marginal effects from regression models. https://doi.org/10.21105/joss.00772 1.

Ludecke, D., Makowski, D., & Waggoner, P. (2020). *performance: Assessment of regression models performance* (Version 0.4.4) https://CRAN.R-project.org/package=performance.

Mackintosh, J. J., & Cauty, A. (1971). Spatial reversal learning in rats, pigeons, and goldfish. *Science, 22*, 281–282.

Madden, J. R., Langley, E. J. G., Whiteside, M. A., Beardsworth, C. E., & van Horik, J. O. (2018). The quick are the dead: Pheasants that are slow to reverse a learned association survive for longer in the wild. *Philosophical Transactions of the Royal Society B: Biological Sciences, 373*. Article 20170297.

Meigs, J. B., Smith, D. C., & Van Buskirk, J. (1983). Age determination of black-capped chickadees. *Journal of Field Ornithology, 54*(3), 283–286.

Mistlberger, R. E. (2009). Food-anticipatory circadian rhythms: Concepts and methods. *European Journal of Neuroscience, 30*, 1718–1729.

Murphy, R. A., Mondragon, E., & Murphy, V. A. (2008). Rule learning by rats. *Science, 319*, 1849–1851. https://doi.org/10.1126/science.1151564

Nakagawa, S., & Cuthill, I. C. (2007). Effect size, confidence interval and statistical significance: A practical guide for biologists. *Biological Reviews, 82*, 591–605.

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution, 4*, 133–142. https://doi.org/10.1111/j.2041-210x.2012.00261.x

Pavlov, I. P. (1927). *Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex*. Oxford University Press.

Pitera, A. M., Branch, C. L., Bridge, E. S., & Pravosudov, V. V. (2018). Daily foraging routines in food-caching mountain chickadees are associated with variation in environmental harshness. *Animal Behaviour, 143*, 93–104. https://doi.org/10.1016/j.anbehav.2018.07.011

Pravosudov, V. V., & Clayton, N. S. (2001). Effects of demanding foraging conditions on cache retrieval accuracy in food-caching mountain chickadees *(Poecile gambeli)*. In *Proceedings of the Royal Society of London, Series B: Biological Sciences, 268* pp. 363–368).

Pravosudov, V. V., & Clayton, N. S. (2002). A test of the adaptive specialization hypothesis: Population differences in caching, memory, and the hippocampus in black-capped chickadees *(Poecile atricapilla)*. *Behavioral Neuroscience, 4*, 512–522.

Pyle, P. (1997). Molt limits in North American passerines. *North American Bird Bander, 22*, 49–89.

R Core Team. (2018). *R: A language and environment for statistical computing* (Version 3.5.2). R Foundation for Statistical Computing https://www.R-project.org/.

Rayburn-Reeves, R. M., Molet, M., & Zentall, T. R. (2011). Simultaneous discrimination reversal learning in pigeons and humans: Anticipatory and perseverative errors. *Learning & Behavior, 39*(2), 125–137. https://doi.org/10.3758/s13420-010-0011-5

Roberts, W. A. (2002). Are animals stuck in time? *Psychological Bulletin, 128*(3), 473–489. https://doi.org/10.1037/0033-2909.128.3.473

Roberts, W. A., & Feeney, M. C. (2009). The comparative study of mental time travel. *Trends in Cognitive Sciences, 13*(6), 271–277. https://doi.org/10.1016/j.tics.2009.03.003

Roberts, W. A., Feeney, M. C., MacPherson, K., Petter, M., McMillan, N., & Musolino, E. (2008). Episodic-like memory in rats: Is it based on when or how long ago? *Science, 320*, 113–115. https://doi.org/10.1126/science.1152709

Rodriguez, W. A., Borbely, L. S., & Garcia, R. S. (1993). Attenuation by contextual cues of retroactive interference of a conditional discrimination in rats. *Animal Learning & Behavior, 21*, 101–105.

Schielzeth, H., Dingemanse, N. J., Nakagawa, S., Westneat, D. F., Allegue, H., Teplitsky, C., Reale, D., Dochtermann, N. A., Garamszegi, L. Z., & Araya-Ajoy, Y. G. (2020). Robustness of linear mixed-effects models to violations of distributional assumptions. *Methods in Ecology and Evolution, 11*(9), 1141–1152. https://doi.org/10.1111/2041-210X.13434

Schielzeth, H., & Forstmeier, W. (2009). Conclusions beyond support: Overconfident estimates in mixed models. *Behavioral Ecology, 20*(2), 416–420. https://doi.org/10.1093/beheco/arn145

Seu, E., Lang, A., Rivera, R. J., & Jentsch, J. D. (2009). Inhibition of the norepinephrine transporter improves behavioral flexibility in rats and monkeys. *Psychopharmacology, 202*, 505–519. https://doi.org/10.1007/s00213-008-1250-4

Shettleworth, S. J. (2010). *Cognition, evolution, and behavior* (2nd ed.). Oxford University Press.

Shimp, C. P. (1976). Short-term memory in the pigeon: Relative recency. *Journal of the Experimental Analysis of Behavior, 25*, 55–61.

Shimp, C. P., & Moffitt, M. (1974). Short-term memory in the pigeon: Stimulus–response associations. *Journal of the Experimental Analysis of Behavior, 22*, 507–512.

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

Stagner, J. P., Michler, D. M., Rayburn-Reeves, R. M., Laude, J. R., & Zentall, T. R. (2013). Midsession reversal learning: Why do pigeons anticipate and perseverate? *Learning & Behavior, 41*, 54–60. https://doi.org/10.3758/s13420-012-0077-3

Stanton, L. A., Bridge, E. S., Huizinga, J., Johnson, S. R., Young, J. K., & Benson-Amram, S. (2021). Variation in reversal learning by three generalist mesocarnivores. *Animal Cognition, 24*, 555–568. https://doi.org/10.1007/s10071-020-01438-4

Stephens, D. W. (1991). Change, regularity, and value in the evolution of animal learning. *Behavioral Ecology, 2*, 77–89. https://doi.org/10.1093/beheco/2.1.77

Storm, B. C., & Bjork, R. A. (2016). Do learners predict a shift from recency to primacy with delay? *Memory & Cognition, 44*, 1204–1214.

Strang, C. G., & Sherry, D. F. (2014). Serial reversal learning in bumblebees (*Bombus impatiens*). *Animal Cognition, 17*(3), 723–734. https://doi.org/10.1007/s10071-013-07041

Suddendorf, T., & Corbalis, M. C. (2007). The evolution of foresight: What is mental time travel, and is it unique to humans? *Behavioral and Brain Sciences, 30*, 299–351. https://doi.org/10.1017/S0140525X07001975

Tello-Ramos, M. C., Branch, C. L., Kozlovsky, D. Y., Pitera, A. M., & Pravosudov, V. V. (2019). Spatial memory and cognitive flexibility trade-offs: To be or not to be flexible, that is the question. *Animal Behaviour, 147*, 129–136. https://doi.org/10.1016/j.anbehav.2018.02.019

Tello-Ramos, M. C., Branch, C. L., Pitera, A. M., Kozlovsky, D. Y., & Pravosudov, V. V. (2018). Memory in wild mountain chickadees from different elevations: Comparing first-year birds with older survivors. *Animal Behaviour, 137*, 149–160. https://doi.org/10.1016/j.anbehav.2017.12.019

Tille, Y., Newman, J. A., & Healy, S. D. (1996). New tests for departures from random behavior in spatial memory experiments. *Animal Learning & Behavior, 24*, 327–340.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology/Psychologie Canadienne, 26*, 1–12.

van Horik, J. O., Beardsworth, C. E., Laker, P. R., Langley, E. J. G., Whiteside, M. A., & Madden, J. R. (2019). Unpredictable environments enhance inhibitory control in pheasants. *Animal Cognition, 22*, 1105–1114.

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag.

Wixted, J. T. (2004). The psychology and neuroscience of forgetting. *Annual Review of Psychology, 55*, 235–269.

Appendix

Singular Fit Justification and Discussion

In this study, we present a generalized linear mixed-effects model (GLMM) with a singular fit and use the results for interpretation. The singular fit warning from the R package ‘lme4’ indicates that ‘some “dimensions” of the variance–covariance matrix have been estimated as exactly zero’ (Bates et al., 2015). Here, we include a brief discussion of why we used this model despite the warning and some potential pitfalls.

The singular fit model (Table 4) and two associated post hoc GLMM models were used as part of a confirmatory hypothesis testing approach. We expected the probability that a chickadee made a reversal error during the first trial of each day to vary with reversal learning ability across days in the first serial reversal experiment (Serial R1). Thus, we fitted a GLMM (binomial distribution, logit link, bobyqa optimizer) with the following focal fixed effects: day, single reversal learning ability (determined from performance on a previous cognitive task) and an interaction of day with single reversal learning ability. We added the following fixed effects to control for other variation in the data: the number of trials completed the previous day and an interaction between day and elevation. Most importantly for this discussion, we fitted a random intercept of individual bird to control for pseudoreplication, as the data set comprised 97 birds observed six times (once per day).

There are several potential causes of a singular fit warning from ‘lme4’. The model may be overfitted from including too many variables, or the random effects may be improperly specified. Importantly, the authors of ‘lme4’ specifically allowed the software to report singular fit models because these models can still provide insight into the trends of the data (Bates et al., 2015). To explore this, we used a stepwise approach to remove variables from the model and evaluate the variance of the random effects.

We determined that the model was not overfit because of too many parameters (i.e. including interaction effects of day∗elevation and the fixed effect of the total trials completed the previous day). Instead, the warning appeared to be the result of covariance between the random intercept (individual bird) and the cognitive variable we included (single reversal learning ability). In other words, the variation in the response variable (whether or not the bird made a reversal error in the first trial of each day) that was explained by the random intercept (individual bird) was almost entirely explained by the fixed effect of single reversal learning ability, resulting in a zero variance estimate for the random effect. However, we could not remove the random effect, due to pseudoreplication issues, and we could not remove the focal fixed effect of single reversal learning ability. Furthermore, this singular fit is harmonious with our overall hypotheses: we expected that better performance on a single reversal learning task would correlate with lower odds of visiting the previous day's rewarding feeder on a serial reversal learning task, suggesting that birds were better able to switch to the new feeder without relying on memories of the previous feeder location.

Finally, there are several studies that suggest linear mixed-effects models are highly robust and that many issues with assumptions and fitting can be resolved through analysing the results within an ecological context (Schielzeth et al., 2020). Thus, we have chosen to include this singular fit model and additionally include the variance of the random intercepts in the main text so the reader can assess the data and see the trends, in addition to the shortcomings of the model. As the purpose of this study was to explore specific hypotheses and compare fixed effects, we chose to fit linear models to approximate differences in learning between individuals, rather than fit nonlinear learning curves to best describe learning on the serial reversal tasks. Thus, we feel confident reporting this singular fit model and the two singular fit post hoc models associated with it.

There are several possible pitfalls to using this model for interpretation. First, models with random intercepts (such as the singular fit model we report) have been shown to increase type I error in comparison to models fitted with random slopes (e.g. Schielzeth & Forstmeier, 2009). But we did not reformat our random effects structure because the random intercept more appropriately fit our a priori hypotheses, and because one approach to reduce singular fit is to simplify the random effects. This usually involves removing random slope effects first, which we had already done in our models. Second, another approach is to ‘keep it maximal’, or keep as many effects as possible without leading to a singular fit (Barr et al., 2013). One advantage of this approach is maximally fit models often perform strongly and have strong power. Thus, although we could not remove the variable that produced the singular fit, we kept as many variables as we felt appropriate to support the hypothesis in question. Finally, in general, the key issue with singular fit models (also called overfitting) is that while they tightly describe the sample of data, they generally do not predict future data and trends as well. This suggests that interpretations of this model might not apply to the overall population, only to our sample. In our case, the model suggests that birds with worse single reversal learning ability were more likely to visit the previous feeder in the first trial each day of Serial R1, and that this effect varied between elevations. We also present a similar model with a different cognitive ability, and so we report that the noncognitive fixed effects in both models are highly consistent. This gives us additional confidence that the singular fit model likely also appropriately estimates the effect of reversal learning ability. We feel that this approach balances the risk of a singular fit model appropriately and thus we include it in the main manuscript.

Table A1
Excluded birds by elevation and serial reversal task

Data set	Serial R1			Serial R2			Combined Serial R1 + Serial R2		
	High	Low	Total	High	Low	Total	High	Low	Total
Before exclusions	168	104	272	194	125	319	196	125	321
After exclusions	51	47	98	149	86	235	154	92	246
Number excluded	117	57	174	45	39	84	42	33	75

Table A2
Models using trial 1 performance from all 7 days of Serial R1

Focal variable	Effect	Location errors			Log-odds visiting previous feeder		
		Estimate	χ^2_1	*P*	Estimate	χ^2_1	*P*
Spatial learning and memory	Day	−0.26	42.30	**<0.001**	−0.59	10.16	**0.001**
	Elevation	—	12.92	**<0.001**	—	0.001	0.98
	Spatial learning and memory	0.07	3.30	0.07	0.11	0.68	0.41
	Day∗elevation	0.19	14.04	**<0.001**	0.70	7.88	**0.005**
		R^2c = 0.28, R^2m = 0.11			R^2c = 0.09, R^2m = 0.05		
Single reversal ability	Day	−0.26	42.14	**<0.001**	−0.62	10.64	**0.001**
	Elevation	—	15.97	**<0.001**	—	1.04	0.31
	Single reversal	0.23	31.00	**<0.001**	0.84	15.93	**<0.001**
	Day∗elevation	0.19	13.99	**<0.001**	0.73	8.33	**0.004**
	Single reversal∗elevation	−0.18	6.06	**0.01**	−1.27	18.57	**<0.001**
		R^2c = 0.29, R^2m = 0.18			Singular fit		

Models do not control for the number of trials completed the previous day (one model for each cognitive task). Conditional and marginal R^2 (R^2c and R^2m) were calculated according to Nakagawa and Schielzeth (2013). Unstandardized regression coefficients used as estimates. Bold text indicates *P* values less than the significance level ($\alpha = 0.05$).

Table A3
Models for Serial R1 with age as a predictor, included as categorical age (juvenile versus adult) or numeric age (years)

Focal variable	Effect	Location errors in trial 1			Location errors in trials 2–5			Log-odds of visiting the previous feeder in trial 1			Number of errors before visiting previous trial 1		
		Estimate	χ^2_1	*P*	Estimate	χ^2	*P*	Estimate	χ^2_1	*P*	Estimate	χ^2_1	*P*
Age (as categorical)	Day	−0.18	16.30	**0.001**	−0.01	0.03	0.87	−0.50	7.04	**0.008**	−0.21	0.32	0.57
	Elevation	—	12.49	**0.001**	—	0.07	0.79	—	0.11	0.74	—	0.60	0.44
	Age (juvenile or adult)	—	0.05	0.82	—	0.56	0.45	—	1.01	0.31	—	0.74	0.39
	Trials completed previous day	0.01	0.03	0.87	−0.23	5.01	**0.03**	0.74	8.19	**0.004**	−1.31	6.64	**0.01**
	Trial	—	—	—	—	7.25[a]	**0.007**	—	—	—	—	—	—
	Day∗elevation	0.16	7.58	**0.006**	−0.05	0.33	0.56	0.70	6.86	**0.009**	0.42	0.78	0.38
		R^2c = 0.25, R^2m = 0.07			R^2c = 0.13, R^2m = 0.02			Singular fit			R^2c = 0.41, R^2m = 0.11		
Age (in years)	Day	−0.18	16.36	**0.001**	−0.01	0.03	0.85	−0.51	7.09	**0.008**	−0.20	0.33	0.57
	Elevation	—	13.34	**0.001**	—	0.29	0.59	—	0.002	0.97	—	0.74	0.39
	Age (in years)	−0.03	0.36	0.55	−0.01	0.03	0.85	−0.01	0.01	0.93	0.19	0.34	0.56
	Trials completed previous day	0.01	0.01	0.94	−0.25	5.87	**0.02**	0.66	6.55	**0.01**	−1.36	7.17	**0.007**
	Trial	—	—	—	—	7.25[a]	**0.007**	—	—	—	—	—	—
	Day∗elevation	0.16	7.71	**0.005**	−0.05	0.28	0.60	0.72	7.13	**0.008**	0.44	0.83	0.36
		R^2c = 0.25, R^2m = 0.08			R^2c = 0.13, R^2m = 0.02			R^2c = 0.09, R^2m = 0.09			R^2c = 0.41, R^2m = 0.10		

Conditional and marginal R^2 (R^2c and R^2m) were calculated according to Nakagawa and Schielzeth (2013). Unstandardized regression coefficients used as estimates. All χ^2 statistics use 1 degree of freedom, unless otherwise indicated. Bold text indicates *P* values less than the significance level ($\alpha = 0.05$).

[a] The χ^2 statistic reports 3 degrees of freedom.

Table A4
Neutral errors before making a reversal error in trial 1 each day of Serial R1 and Serial R2

Focal variable	Effect	Serial R1			Serial R2		
		Estimate	χ^2_1	*P*	Estimate	χ^2_1	*P*
Spatial learning and memory	Day	−0.22	0.41	0.52	−0.97	19.73	**<0.001**
	Elevation	—	0.83	0.36	—	7.01	**0.008**
	Spatial learning and memory	0.13	0.29	0.59	0.21	6.00	**0.01**
	Day∗elevation	0.16	0.14	0.71	0.29	1.40	0.24
		R^2c = 0.31, R^2m = 0.02			R^2c = 0.27, R^2m = 0.21		
Single reversal performance	Day	−0.23	0.44	0.51	−1.13	18.94	**<0.001**
	Elevation	—	1.12	0.29	—	7.18	**0.007**
	Single reversal performance	0.39	2.69	0.10	0.03	0.12	0.72
	Day∗elevation	0.17	0.16	0.69	0.45	2.47	0.11
		R^2c = 0.31, R^2m = 0.04			R^2c = 0.30, R^2m = 0.23		

Separate models were fitted for spatial learning and memory and single reversal performance, due to covariance. Conditional and marginal R^2 (R^2c and R^2m) were calculated according to Nakagawa and Schielzeth (2013). Unstandardized regression coefficients used as estimates. Bold text indicates *P* values less than the significance level ($\alpha = 0.05$).

Table A5
Models results for Serial R2 age regressions

Focal variable	Effect	Location errors in trial 1			Location errors in trials 2–5			Log-odds of visiting the previous feeder in trial 1			Number of errors when visiting previous trial 1		
		Estimate	χ^2_1	*P*	Estimate	χ^2	*P*	Estimate	χ^2_1	*P*	Estimate	χ^2_1	*P*
Age (as categorical)	Day	−0.24	131.24	**<0.001**	−0.59	292.63	**<0.001**	−0.06	0.37	0.54	−0.97	21.56	**<0.001**
	Elevation	—	1.05	0.31	—	15.12	**<0.001**	—	5.22	**0.02**	—	5.86	**0.02**
	Age (juvenile or adult)	—	0.61	0.43	—	0.77	0.38	—	0.20	0.65	—	1.04	0.31
	Trials completed previous day	0.08	9.53	**0.002**	0.17	16.49	**<0.001**	0.51	23.53	**<0.001**	−0.12	0.57	0.45
	Trial	—	—	—	—	55.52[a]	**<0.001**	—	—	—	—	—	—
	Day∗elevation	0.08	8.66	**0.003**	0.36	73.89	**<0.001**	0.18	2.76	0.10	0.19	0.66	0.42
		R^2c = 0.16, R^2m = 0.09			R^2c = 0.21, R^2m = 0.10			R^2c = 0.06, R^2m = 0.03			R^2c = 0.30, R^2m = 0.20		
Age (in years)	Day	−0.24	131.52	<0.001	−0.60	292.04	**<0.001**	−0.05	0.36	0.55	−0.97	21.55	**<0.001**
	Elevation	—	1.06	0.30	—	15.41	**<0.001**	—	5.36	**0.02**	—	5.28	**0.02**
	Age (in years)	−0.02	1.26	0.26	−0.05	1.72	0.19	0.04	0.51	0.47	−0.01	0.02	0.87
	Trials completed previous day	0.08	9.27	0.002	0.17	16.19	**<0.001**	0.51	24.20	**<0.001**	−0.10	0.35	0.55
	Trial	—	—	—	—	55.51[a]	**<0.001**	—	—	—	—	—	—
	Day∗elevation	0.08	8.75	0.003	0.36	74.10	**<0.001**	0.18	2.71	0.10	0.20	0.68	0.41
		R^2c = 0.16, R^2m = 0.09			R^2c = 0.21, R^2m = 0.10			R^2c = 0.06, R^2m = 0.03			R^2c = 0.30, R^2m = 0.20		

Models use either categorical age (juvenile versus adult) or numeric age (in years). Conditional and marginal R^2 (R^2c and R^2m) were calculated according to Nakagawa and Schielzeth (2013). Top models include age (categorical) and bottom models include age (numeric). Unstandardized regression coefficients used as estimates. All χ^2 statistics use 1 degree of freedom, unless otherwise indicated. Bold text indicates *P* values less than the significance level ($\alpha = 0.05$).

[a] The χ^2 statistic reports 3 degrees of freedom.

Table A6
Models results for Serial R1 and Serial R2 models with total neutral errors for trial 1 and trials 2–5

Response variable	Effect	Spatial serial reversal 1 (*N* = 97)				Spatial serial reversal 2 (*N* = 180)			
		Estimate	SE	χ^2	*P*	Estimate	SE	χ^2	*P*
Neutral errors in trial 1	Day	−0.24	0.06	16.54	**<0.001**	−0.34	0.03	133.20	**<0.001**
	Elevation	0.48	0.12	0.83	**<0.001**	0.08	0.05	2.03	**0.008**
	Spatial learning and memory	0.08	0.06	0.29	0.17	0.00	0.03	0.01	0.94
	Previous reversal trials	−0.07	0.09	0.66	0.42	0.11	0.04	8.00	**0.01**
	Day∗elevation	0.21	0.08	0.14	**0.006**	0.10	0.04	7.34	**0.007**
	Bird ID (random intercept)	—	—	—	Var. = 0.22	—	—	—	Var. = 0.06
				R^2c = 0.34, R^2m = 0.10				R^2c = 0.24, R^2m = 0.13	
				AIC = 1790.08				AIC = 7694.76	
Neutral errors in trials 2–5	Day	0.00	0.08	0.002	0.97	−0.72	0.05	224.15	**<0.001**
	Elevation	0.26	0.17	2.56	0.11	0.32	0.10	11.17	**<0.001**
	Spatial learning and memory	0.03	0.08	0.16	0.69	0.07	0.04	2.88	0.09
	Previous reversal trials	−0.35	0.12	7.75	**0.005**	0.21	0.06	13.60	**<0.001**
	Trial	—		0.62[a]	0.09	—	—	18.38[a]	**<0.001**
	Day∗elevation	−0.03	0.11	0.09	0.77	0.42	0.06	54.10	**<0.001**
	Bird ID (random intercept)	—	—	—	Var. = 0.33	—	—	—	Var. = 0.22
				R^2c = 0.15, R^2m = 0.03				R^2c = 0.22, R^2m = 0.12	
				AIC = 2094.53				AIC = 10412.60	

Conditional and marginal R^2 (R^2c and R^2m) were calculated according to Nakagawa and Schielzeth (2013). Elevation estimates relative to high elevation. Unstandardized regression coefficients used as estimates. All χ^2 statistics use 1 degree of freedom, unless otherwise indicated. Bold text indicates *P* values less than the significance level ($\alpha = 0.05$).

[a] The χ^2 statistic reports 3 degrees of freedom.

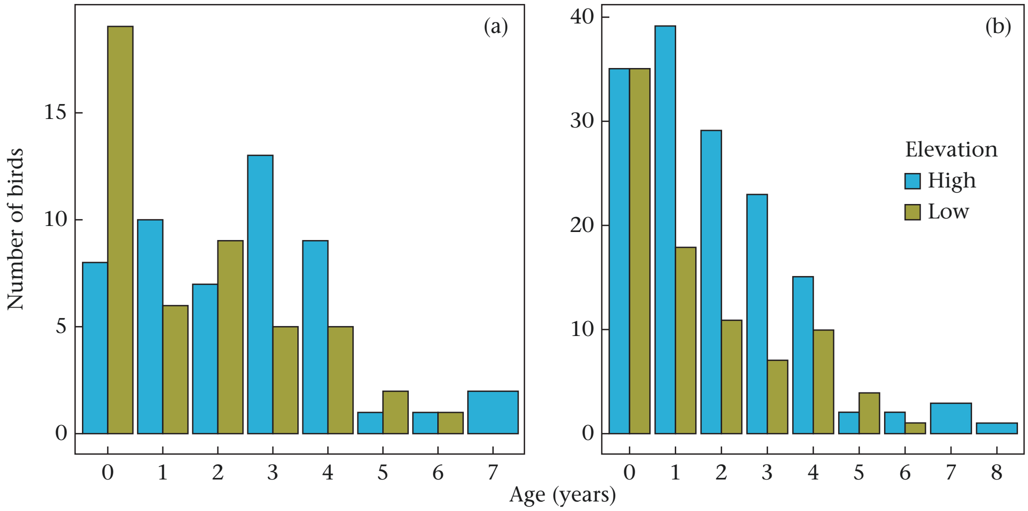

Figure A1. Number of birds by elevation and age in (a) Serial R1 ($N = 97$) and (b) Serial R2 ($N = 235$). Age 0 years indicates juveniles; age 1–8 years indicates adults.

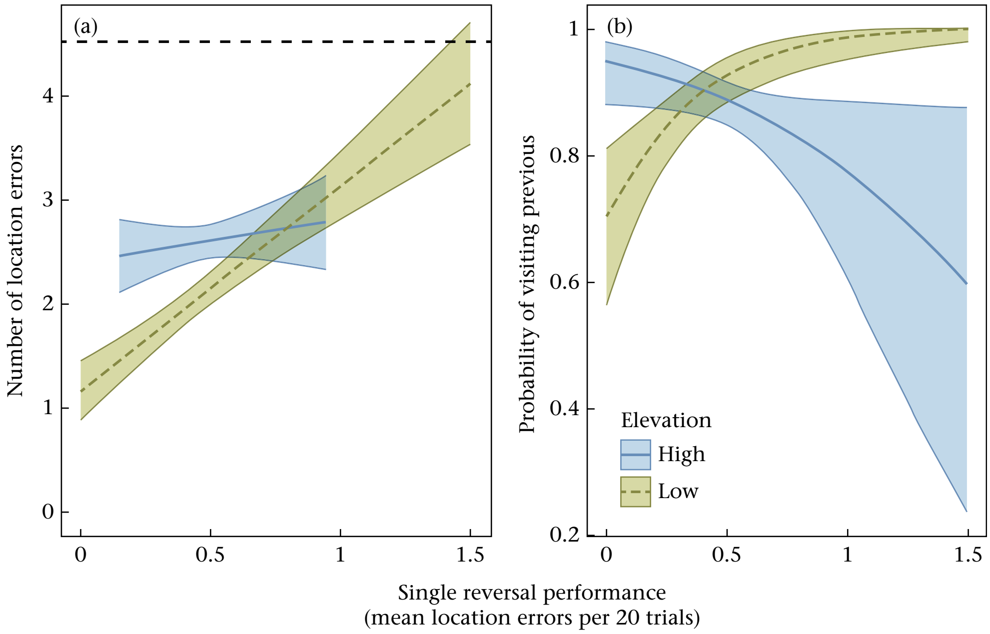

Figure A2. Performance in trial 1 of Serial R1 versus single reversal performance (i.e. cognitive flexibility): (a) location errors; (b) predicted probability of visiting the previous feeder. Heavy dashed line indicates expected random value.

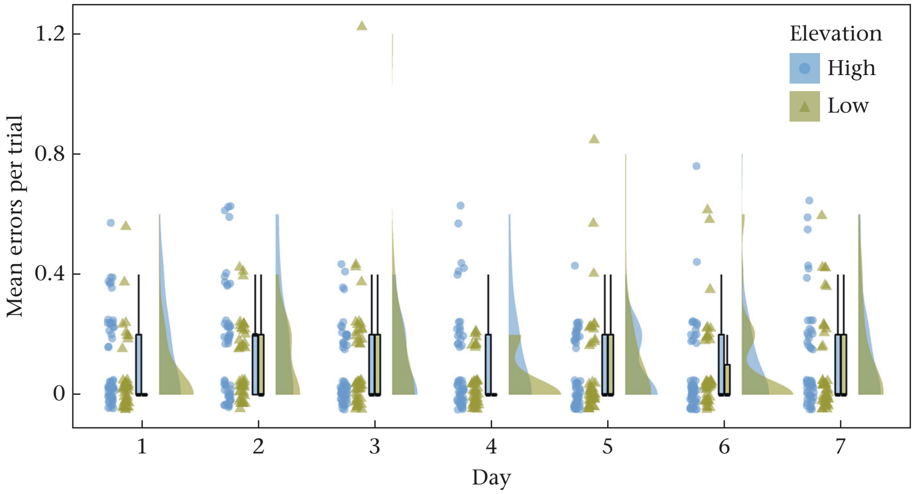

Figure A3. Mean last five errors per trial per day for Serial R1. For box plots, lower and upper box boundaries represent 25th and 75th percentiles, lines inside boxes represent the median, lower and upper error lines indicate 1.5× greater and lesser than the 25th and 75th percentiles, respectively. Outliers not shown. Transparent points (jittered for clearer viewing) and density plots represent raw data.

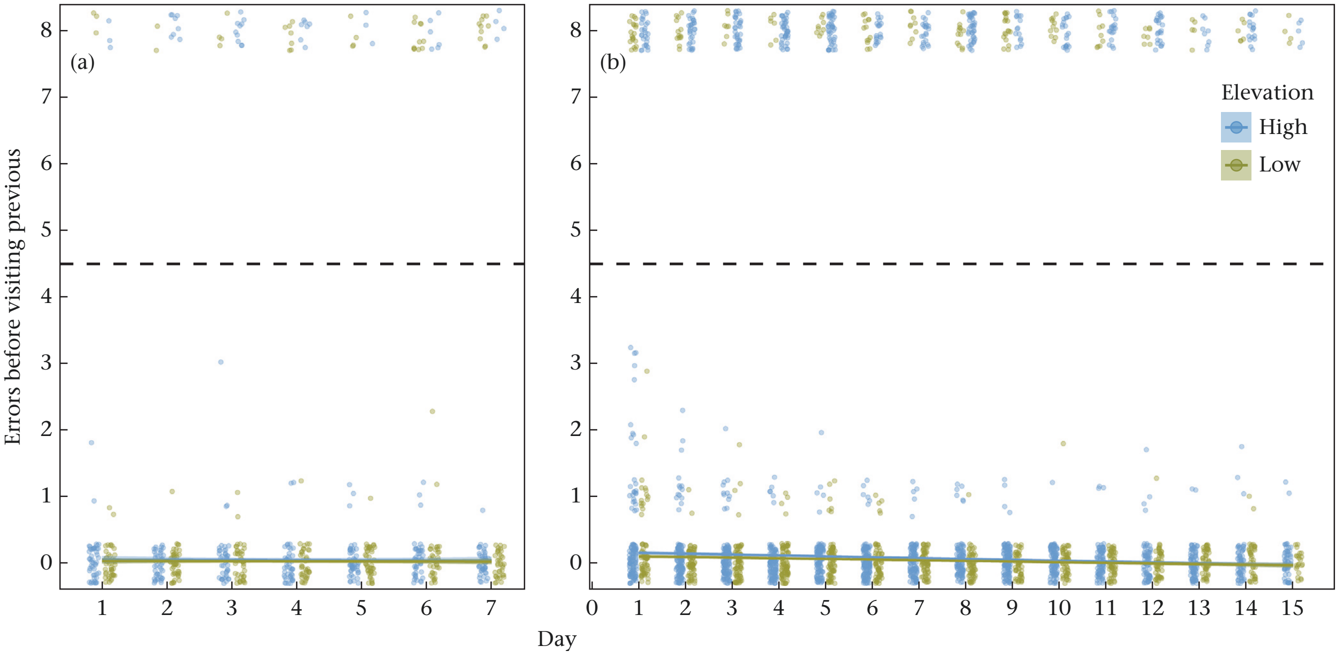

Figure A4. Number of neutral errors before visiting the previous feeder in the first trial per day for (a) Serial R1 ($N = 97$ birds for each day) and (b) Serial R2 ($N = 235$ birds for days 3–10 and 83 birds on day 17). The number of errors before visiting the previous feeder is a subset of the total number of errors a bird made. For example, if a bird made three total errors and two of those errors were before visiting the previous feeder, then this bird visited the previous feeder on its third visit before finding the correct feeder on its fourth visit during the trial. 'NA' indicates that the bird did not visit the previous feeder during the trial, but this does not indicate that these birds made zero errors overall. Dashed lines indicate expected random value (4.5 errors).

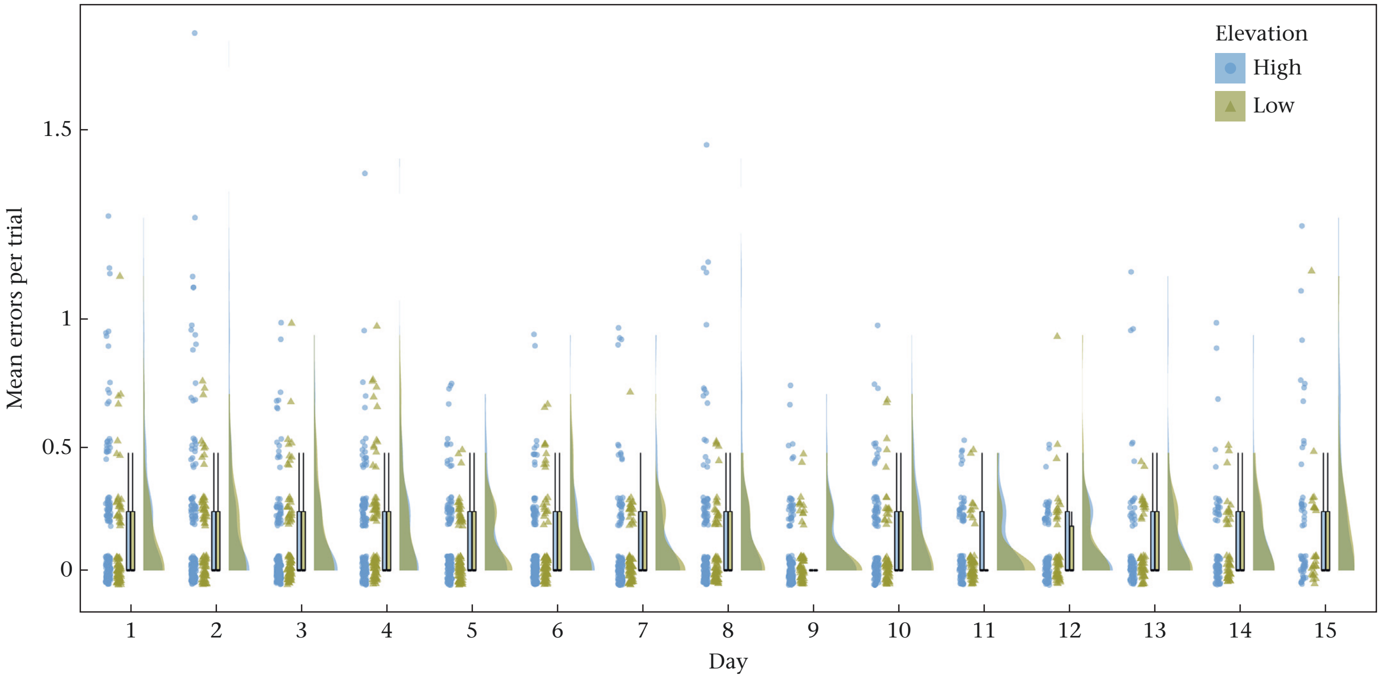

Figure A5. Mean last five errors per trial per day for Serial R2. For box plots, lower and upper box boundaries represent 25th and 75th percentiles, lines inside boxes represent the median, lower and upper error lines indicate 1.5× greater and lesser than the 25th and 75th percentiles, respectively. Outliers not shown. Transparent points (jittered for clearer viewing) and density plots represent raw data.

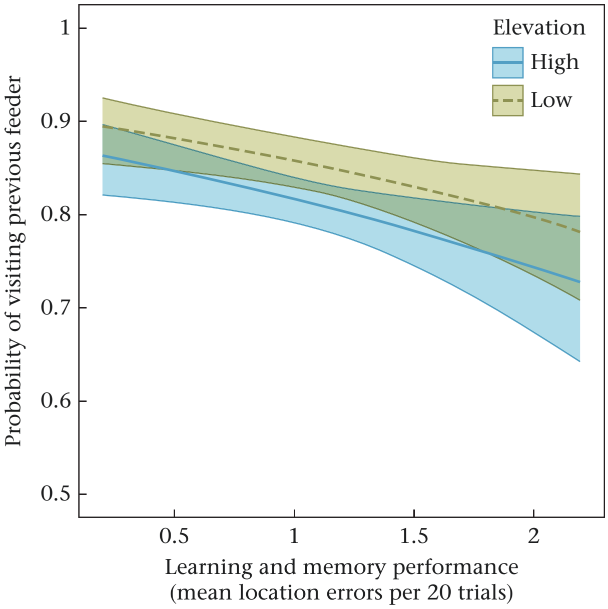

Figure A6. The predicted probability of visiting the previous feeder by spatial learning and memory ability by elevation for Serial R2.