Supplemental Materials

Testing the greater male variability phenomenon: male mountain chickadees exhibit larger variation in reversal learning performance compared to females

Carrie L Branch, Benjamin R Sonnenberg,Angela M Pitera, Lauren M Benedict, Dovid Y Kozlovsky, Eli S Bridge, and Vladimir V Pravosudov

Proceedings of the Royal Society B

DOI: 10.1098/rspb.2020.0895

Supplemental Results

There were slight methodological differences between the 2016 testing year and the other three years of testing (reversal learning task in 2016 commenced after 16 days of spatial learning and memory task as opposed to only 5 days during all other years), therefore, we used a linear model (function *lm()* in base R 3.5.2) to assess differences in performance across years. In models that assessed mean number of errors per trial across the entire testing period, test year and total number of trials completed were included as fixed factors, while number of errors per trial on the first 20 trials only included test year as a fixed factor, as all birds completed the same number of trials. Tukey post-hoc comparisons (*lsmeans()* function in lsmeans package, Lenth 2016) revealed a significant difference in the total number of errors per trial across the entire testing period on the spatial cognitive task between 2017 and 2019 (p = 0.016) and a nearly significant difference between 2016 and 2017 (p = 0.054).

For the reversal cognitive task, testing year 2016 differed significantly from the other 3 years of testing on both response measures; errors per trial on the first 20 trials and across total trials, see Table S2. All other pairwise comparisons between the number of errors on the first 20 trials and overall performance for both cognitive tasks were not significant (p > 0.05). Due to the differences in methodology and performance we ran separate analyses on 2016 alone, all four years together, and on all 3 comparable years without 2016.

Cognitive performance in 2016

*Spatial cognitive task*

Across the first 20 trials, where learning acquisition was tested, there was no significant difference between male (mean errors/trial: 0.978 ± 0.558) and female (0.914 ± 0.515) performance on the spatial learning and memory task (t = 0.478, p = 0.634), nor was there a significant effect of elevation (t = -0.035, p = 0.972, published in Croston et al. 2017). Using Levene’s test for homogeneity of variance, variances between males and females were homogenous (F3,80= 0.715, p = 0.546), meaning there was no difference in the range of variation in performance.

            Across all trials, where overall cognitive performance was tested, there was no significant differences between male (0.238 ± 0.15) and female (0.254 ± 0.19) performance (t = 0.783, p = 0.436), however, there was a significant effect of elevation (t = -2.759, p = 0.007, previously published in Croston et al. 2016). As noted in the methods, the total number of trials was significantly associated with performance (t = -7.615, p < 0.0001). Using Levene’s test for homogeneity of variance, variances between males and females were homogenous (F3,80= 0.297, p = 0.828), indicating no significant difference in the range of variation between males and females.

*Reversal spatial cognitive task*

There was no significant difference between male (mean errors/trial: 0.884 ± 0.676) and female (0.657 ± 0.494) learning acquisition on the reversal spatial cognitive task across the first 20 trials (t = 0.979, p = 0.331). However, birds from high elevation performed significantly worse compared to birds at low elevations (t = -2.781, p = 0.007; as in Croston et al. 2017 and Tello-Ramos et al. 2018). Using Levene’s test for homogeneity of variance, variances between males and females were homogenous (F3,67= 1.323, p = 0.274), indicating no significant difference in the range of variation between males and females.

            There was no significant difference between male (mean errors/trial: 0.372 ± 0.581) and female (0.231 ± 0.419) overall performance across all trials (t = 1.429, p = 0.158) on the reversal learning task; however, there was a significant effect of elevation (t = -4.237, p < 0.0001, Croston et al. 2017). Again, as noted in the methods, the total number of trials was significantly associated with performance (t = -5.435, p < 0.0001). Using Levene’s test for homogeneity of variance, variances between males and females were homogenous (F3,67= 2.274, p = 0.088), meaning there was no significant difference in the range of variation in performance.

Cognitive performance without 2016

Due to slight methodological differences between 2016 and the other 3 years of testing (2017, 2018 and 2019) discussed in the methods, we conducted additional analyses removing the 2016 data. When 2016 was removed, all non-significant effects remained non-significant (p > 0.05) and the significant difference in male and female range of variation on the reversal spatial cognitive task disappeared (see Table S3).

Comparing individual performance on spatial and reversal cognitive tasks

We assessed whether individuals that performed well on the spatial learning and memory task also performed well on the reversal spatial learning task. We ran separate GLMMs (*lmer()* in R version 3.6.1, packages lme4 and lmerTest; Bates et al. 2015 and Kuznetsova et al. 2017) on cognitive performance with males and females run together, males alone, and females alone. For each model, performance on reversal (number of errors on the first 20 trials and across the entire task) was the response variable and performance on the spatial learning and memory task and elevation were fixed effects, along with individual ID and test year as random effects.

These models were run for number of errors on the first 20 trials and across the entire task. For models run across the entire task we also included total number of trials completed on the reversal task as a fixed effect. Only total number of trials on the reversal learning and memory task was included in the model, as the number of trials completed on the spatial learning and memory task and the reversal task were highly correlated (Trials completed: F1, 139.93 = 39.74, p < 0.0001; Elevation: F1, 116.75 = 10.96, p = 0.001).

*Spatial performance does not predict reversal performance on the first 20 trials*

When males and females were run together, performance on the spatial cognitive task did not significantly predict performance on the reversal cognitive task, however, elevation did significantly predict performance on the reversal task (Performance: F1, 204.03 = 0.005, p = 0.946; Elevation: F1, 116.75 = 15.67, p = 0.0001). We found the same pattern when males were run alone (Performance: F1, 115.31 = 0.324, p = 0.571; Elevation: F1, 118 = 5.65, p = 0.019) and when females were run alone (Performance: F1, 82.95 = 0.513, p = 0.476; Elevation: F1, 83.39 = 8.61, p = 0.004).

*Spatial performance predicts reversal performance* *across the entire testing period*

When males and females were run together, performance on the spatial cognitive task approached, but did not reach statistical significance at p < 0.05 in predicting performance on the reversal cognitive task; however, total trials completed and elevation did significantly predict performance on the reversal task (Performance: F1, 203.49 = 3.54, p = 0.062; Total trials: F1, 205.60 = 54.37, p < 0.0001; Elevation: F1, 205.94 = 23.80, p < 0.0001; Figure S1a).

When males were run alone, performance on the spatial cognitive task did not significantly predict performance on the reversal cognitive task, however, total trials completed and elevation were significantly associated with performance on the reversal task (Performance: F1, 115.08 = 0.375, p = 0.542; Total trials: F1, 116.98 = 39.77, p < 0.0001; Elevation: F1, 116.52 = 16.79, p < 0.0001; Figure S1b).

When females were run alone, performance on the spatial cognitive task significantly predicted performance on the reversal task as did total trials completed, however, elevation was not significantly associated with reversal performance (Performance: F1, 80.19 = 7.80, p = 0.007; Total trials: F1, 45.24 = 14.98, p = 0.0003; Elevation: F1, 49.80 = 3.32, p = 0.074; Figure S1c).

Supplemental Figure

Figure S1. Mean number of errors over the entire task on the spatial learning and memory task regressed with the mean number of errors on the reversal spatial task: a). males and females run together, b). males run alone and c). run females alone. For all statistical models reported in the results, elevation predicted mean number of errors on the reversal spatial learning task. High (H) elevation individuals are represented by grey circles and low (L) elevation individuals are represented by black triangles. Lines and standard error ranges are a result of linear models run for the mean number of errors on spatial learning and memory task by the mean number of errors on the reversal spatial learning task. Individuals did not have negative mean number of errors on either task; individual points have been offset using *geom\_jitter()* in R package ggplot2 (Wickham 2016).



Figure S1a.



Figure S1b.



Figure S1c.

Supplemental Tables

Table S1. Sample sizes across four years of testing at High and Low elevation sites for spatial learning and memory task and reversal spatial cognitive tasks.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **High** |  | **Low** |  |
| Test Year | Spatial | Reversal | Spatial | Reversal |
| 2016 | 43 | 34 | 41 | 37 |
| 2017 | 62 | 56 | 51 | 10 |
| 2018 | 52 | 48 | 0 | 0 |
| 2019 | 44 | 20 | 8 | 5 |
| Total | 201 | 158 | 100 | 52 |

Table S2. Results of pairwise Tukey post-hoc comparison across all 4 years of testing. P-values significant at p < 0.05 are bolded. Values in italics approach significance.

|  |  |  |
| --- | --- | --- |
|  |  |  |
| Trial Type | Comparison | p-value |
| Spatial |  |  |
| First 20 trials | 2016 to 2017 | 0.999 |
|  | 2016 to 2018 | 0.989 |
|  | 2016 to 2019 | 0.890 |
|  | 2017 to 2018 | 0.976 |
|  | 2017 to 2019 | 0.907 |
|  | 2018 to 2019 | 0.787 |
|  |  |  |
| Total trials | *2016 to 2017* | *0.054* |
|  | 2016 to 2018 | 0.573 |
|  | 2016 to 2019 | 0.999 |
|  | 2017 to 2018 | 0.572 |
|  | **2017 to 2019** | **0.016** |
|  | 2018 to 2019 | 0.464 |
|  |  |  |
| Reversal |  |  |
| First 20 trials | **2016 to 2017** | **0.008** |
|  | **2016 to 2018** | **0.004** |
|  | **2016 to 2019** | **0.0004** |
|  | 2017 to 2018 | 0.973 |
|  | 2017 to 2019 | 0.357 |
|  | 2018 to 2019 | 0.607 |
|  |  |  |
| Total trials | **2016 to 2017** | **0.0001** |
|  | **2016 to 2018** | **0.0002** |
|  | **2016 to 2019** | **0.0007** |
|  | 2017 to 2018 | 0.996 |
|  | 2017 to 2019 | 0.948 |
|  | 2018 to 2019 | 0.986 |

Table S3. Results of analyses run without the 2016 testing year, including years 2017 – 2019. Male and female performance on the spatial and reversal cognitive tasks (mean number of errors and standard deviation). The results of GLMMs and homogeneity of variance tests are also reported. GLMMs were run exactly as reported in the methods for all 4 years.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
| Trial Type | Male | Female | GLMM |  |  |  |
| Spatial | N = 110 | N = 107 | Sex | Elevation | Total Trials | Hom Var |
| First 20 trials | 0.929 ± 0.417 | 0.968 ± 0.42 | t = -0.343, p = 0.732 | t = 0.461, p = 0.646 | NA | F = 1.849, p = 0.175 |
| Total testing | 0.366 ± 0.209 | 0.386 ± 0.248 | t = -0.204, p = 0.839 | t = 0.914, p = 0.362 | t = -10.131, p = 0.0001 | F = 1.585, p = 0.209 |
|  |  |  |  |  |  |  |
| Reversal | N = 73 | N = 66 | Sex | Elevation |  | Hom Var |
| First 20 trials | 0.547 ± 0.229 | 0.537 ± 0.243 | t = -0.023, p = 0.982 | t = -1.871, p = 0.064 | NA | F = 0.0001, p = 0.991 |
| Total testing | 0.25 ± 0.139 | 0.127 ± 0.270 | t = -0.093, p = 0.926 | t = -1.608, p = 0.110 | t = -6.233, p = 0.0001 | F = 0.034, p = 0.854 |