**Supplemental Information**

**An Instance-Based Model Account of the Benefits of Varied Practice in Motor Learning**

Thomas E. Gorman & Robert L. Goldstone

**Supplementary Analyses**

**Experiment 2 Learning Rates and Performance Across Testing Stage**

The testing phase performances reported in the primary manuscript were aggregated over the entirety of the testing phase. However, participants did improve throughout the testing phase, and it is thus possible that the superior performance of the varied group was thus driven by trials at the beginning of testing, as might have occurred if the constant group was disproportionately burdened by switching to throwing positions outside of their training range. Therefore, we conducted additional analyses to address whether the aggregate varied advantage in the testing phase was driven by the very beginning of the testing phase and whether the constant group was catching up to the varied group during testing.

We first compared performance between the first half and second half of testing. Although participants in both conditions completed the same number of testing trials for each testing position, the order in which they encountered each position was randomized. Additionally, because of trial exclusions (e.g. failing to release the ball within the valid zone) participants differed in the number of valid trials included in the testing analysis (although such differences were minimized due to our exclusion criteria). Thus, we indexed each testing trial separately for each participant, and each participant had a separate index for each testing position. For each participant, we then computed the average absolute deviation from the target separately for the first and second half of training (Supplementary Figure 1, top row). Finally, we subtracted the first half performances from second half performances to yield an improvement score for each position (Supplementary Figure 1, bottom-left panel). We collapsed performance scores across position for subsequent analyses, but we note that including position as a within-subjects factor does not influence these results (and there is no significant interaction between position and condition). To assess whether training condition influenced the amount of improvement between the first and second half of testing, we performed type III ANOVA with first half to second half improvement as a dependent measure, and training condition as a between groups factor. This comparison yielded no significant effect of condition, F(1, 206)=.818, p=.367, η2*G* = <.01. Next, we compared performance between conditions using data only from the second half of testing. In this case, the effect of condition was significant F(1, 206)=4.685, p=.032, η2*G* = .02. As was the case in the analyses of the primary manuscript, participants in the varied condition (Mean=189.32, sd=75.12) performed better than participants trained in the constant condition (Mean=222.94, sd=81.58).

The analyses above suggest that training condition did not influence improvement between the first and second half of testing, and that varied participants performed better overall even when only the latter half of testing data is considered. However, it could be argued that our method of binning the data was somewhat arbitrary and may still be obscuring a difference between conditions in rate of improvement. Moreover, previous research in the perceptual learning literature has shown that comparing aggregate testing performance may sometimes obscure differences in learning rate (Kattner et al., 2017). Therefore, to ensure that group differences in learning rate during testing were not influencing the results of the present study, we’ve included a comparison of the rate at which participants in each condition improved during testing.

Following previous research on the functional form of individual learning curves (Heathcote et al., 2000), we fit exponential learning models to the testing performance data for each individual participant. Models were fit using the TEfit package in R (Cochrane, 2020). The exact formula is shown in supplementary Equation 1. The models simultaneously fit 3 parameters corresponding to starting performance, asymptotic performance, and learning rate. Each participant had trial number indices assigned separately to each valid throw for each testing position. For a given testing position, performance was modelled as an exponential function of the trial number for that position. The function produced inverse learning rates, with smaller values corresponding to faster rates of improvement. The fit learning rates, aggregated within each condition, are shown in the bottom-right panel of Supplementary Figure 1. We averaged learning rates across positions and performed a between groups comparison of condition (constant vs. varied) on learning rate. The effect of condition was not significant, F(1,206)=.227 p=.63, η2*G* <.01.

**Supplementary Equation 1:** Asymptotic Performance + (Start Performance-Asymptotic Performance) \*

It should be noted that the testing performance data on the level of individual participants was fairly noisy, and for some participants the model fitting process failed to converge. However, excluding cases where the model did not converge on best fitting parameter values does not influence the pattern of results. The results presented are also consistent with analyses that do not average across position, and there was no interaction between position and condition on learning rate. Fitting the learning models to trial-level data that has first been averaged within each group also produces results consistent with those presented here.

**Supplementary Figure 1**

Chart, bar chart

Description automatically generated

*Supplementary Figure 1: Comparisons of learning rate and staged testing performance for Experiment 2. Top row: left side shows testing performance for each position in the first half of testing, right side shows performance for the 2nd half of testing. Bottom row, left side shows the difference in performance between the 1st and 2nd half of testing. Differences are computed per participant, per position, and then averaged within each condition. Larger differences indicate greater amounts of improvement. Bottom row, right side shows inverse learning rates, fit to each participant for each testing position. Smaller learning rates indicate faster improvement.*

**Interpolation vs. Extrapolation Account**

To address the possibility that the observed effect of training condition was driven exclusively by extrapolation/interpolation effects, we include an additional analysis that explicitly controls for whether testing locations were interpolations or extrapolations from trained location(s). We considered throws from the trained location to be interpolation, and thus each constant group had a single interpolation location during testing, and 5 extrapolation locations. For the varied group, both training positions (500, 800) were considered interpolation, along with the intermediate positions (625, 675). The varied group was then left with 2 extrapolation locations (400, 900). An ANOVA analysis with condition as a between groups factor (constant vs. varied), and location type (extrapolation vs. interpolation) as a within groups factor, revealed a significant effect of training condition F(1,206)=3.94, p=.048, η2*G* = .017. The effect of location type was not significant F(1,206)=.024, p=876, η2*G* = <.001, nor was the interaction between location type and condition F(1,206)=.121, p=.729, η2*G* <.001. Descriptive statistics are displayed in supplemental Table S1.

**Supplemental Table S1**

Table

Description automatically generated with medium confidence

*Supplemental Table 1: Testing performance for varied and constant groups, split by whether testing locations were extrapolations or interpolations from training location(s). Mean absolute deviation from the center of the target, with standard deviations in parenthesis.*

**Analyses with limited participant and trial exclusions**

A considerable amount of data was excluded from the primary analyses in both experiments. Participants were removed from the final data set for three different reasons:

1) reporting using an input device other than a mouse or a trackpad (8 exclusions in experiment 1; 25 in experiment 2).

2) Behavioral patterns indicative of bad faith participation, such as dropping the ball straight down, or repeatedly throwing the ball into the barrier to end trials quickly (16 exclusions in experiment 1: 64 in experiment 2).

3) Statistically significant poor performance compared to other participants in the same condition (13 exclusions in experiment 1; 22 in experiment 2).

We also removed trials in which the ball struck the barrier, and trials with an outlying absolute deviation from the target. The task was calibrated such that a small slip of the mouse could result in the ball traveling a much greater distance (e.g., over 1500 units) beyond the center of the target, far beyond the deviance of the average throw in the early stage of training and we thus believe that many of these throws were reflective of a process unrelated to the general accuracy we sought to measure.

To assess whether our exclusion criteria exert a measurable impact on the general pattern of our behavioral results, we also include a testing performance analysis on a dataset with a far more liberal exclusion criterion. For these analyses, we retain all participants previously excluded for reporting a disallowed input device, and for having poor performance. Data from participants excluded for bad faith behavioral patterns were not included, as their data does not appear to reflect accuracy at the task. These relaxed exclusion analyses also retain the previously excluded barrier-hit trails.

For Experiment 1 a total of 86 participants were retained for the relaxed exclusion criterion analysis (n=42 varied; n=44 constant). We performed an ANOVA comparing training condition (constant vs. varied), with throwing position as a within-participants factor, and participant inclusion status (included vs. excluded) as a covariate. Significant effects were observed for condition F(1, 82)=4.13, p =.045, η2*G* = .021, inclusion status F(1, 82)=55.68, p <.001, η2*G* = .284, and throwing position F(3, 246)=44.72, p <.001, η2*G* = .128. There was no significant interaction between condition and inclusion status, F(1, 82<.001, p =.99, η2*G* <.001. Supplementary Figure 2 displays the testing performance of the augmented Experiment 1 dataset, and separate performances for included and excluded participants.

For Experiment 2 a total of 230 participants were retained for the relaxed exclusion criterion analysis (n=36 varied; n=194 constant). Again, we performed an ANOVA comparing training condition (constant vs. varied), with participant inclusion status in the primary analysis (included vs. excluded) and throwing position as covariates. In this case, the effect of condition did not reach significance F(1, 226)=2.91, p =.089, η2*G* = .007. Significant effects were obtained for Inclusion status F(1, 226)=57.12, p <.001, η2*G* = .284, and throwing position F(5, 1130)=157.96, p <.001, η2*G* = .128. There was no significant interaction between condition and inclusion status, F(1, 226)=.002, p =.96, η2*G* <.001. Supplementary Figure 2B displays the testing performance of the augmented experiment 2 dataset, and separate performances for included and excluded participants.

**Supplementary Figure 2A**

Chart, bar chart

Description automatically generated

**Supplementary Figure 2B**

Chart, bar chart

Description automatically generated

*Supplementary Figure 2. Performance at each testing location. The top row shows performance with the enlarged datasets, bottom row separates performance for participants that were excluded or included from primary analyses. Figure 2A and 2B correspond to Experiment 1 and Experiment 2, respectively. Error bars indicate standard error of the mean.*

*Signed deviation performance measure*

**Supplemental Table S2.** Signed deviation measure of performance. Positive values reflect shooting beyond the target center. Tables present mean distance from target with standard deviation in parentheses.

**Experiment 1**

A picture containing text

Description automatically generated

**Experiment 2**

A picture containing text

Description automatically generated

**Supplemental Figure 3A**

Chart, bar chart, waterfall chart

Description automatically generated

**Supplemental Figure 3B**

Chart, bar chart

Description automatically generated

*Supplemental Figure 3: Throws are split into overthrows which went beyond the target and underthrows which fell short. 3A shows performance from the testing phase of experiment 1, 3B shows performance from the testing phase of Experiment 2. Error bars indicate standard error of the mean.*

**Experiment 1 Intermittent Testing**

In experiment 1, participants in both conditions completed brief intermittent testing stages following every 20 trials of training. For both conditions, each intermittent testing stage consisted of 2 trials from positions 610, 760 and 910, presented in randomized order. No feedback was presented on these trials. After subjects released the ball, the ball disappeared from the screen, and participants were prompted to begin the next trial without receiving any information concerning how far away from the target their throw would have landed. Supplemental table 3, and Supplemental Figure 4 present data aggregated from the first half of the intermittent testing stages (i.e. stages 1-5), and the final half (i.e. stages 6-10). Intermittent testing was not included in experiment 2.

**Supplemental Table S3**

**A picture containing table

Description automatically generated**

*Supplemental Table S3:* Absolute deviation measure of performance from the intermittent testing stages of experiment 1. Tables present mean distance from target with standard deviation in parentheses.

**Supplemental Figure 4**

**Chart, bar chart

Description automatically generated**

*Supplemental Figure 4: Intermittent testing performance, binned into the first half and second half.. Error bars indicate standard error of the mean.*