Practice schedules affect how learners correct their errors: Secondary analysis from a contextual interference study.

Taylor, S.,[1] Fawver, B.,[2] Thomas, J.L.,[1] Williams, A.M., [1,3] & Lohse, K.R.[4]

1 Department of Health and Kinesiology, University of Utah, Salt Lake City, UT

2 US Army Medical Research Directorate-West, Walter Reed Army Institute of Research

3 Institute for Human and Machine Cognition, Pensacola, FL

4 Physical Therapy and Neurology, Washington University School of Medicine, Saint Louis, MO

PLEASE NOTE THAT THIS IS A PRE-PRINT AND HAS NOT BEEN PEER-REVIEWED.
Please cite as: Taylor, S., Fawver, B., Thomas, J.L., Williams, A.M., & Lohse, K.R. (2022). Practice schedules affect how learners correct their errors: Secondary analysis from a contextual interference study. *SportRxiv*.

Date Submitted: April 4th, 2022

Keywords: errors; phase space; random practice; timing;

Author(s) Conflict of Interest and Disclosures: The authors report no conflict of interest or disclosures. Material has been reviewed by the Walter Reed Army Institute of Research. There is no objection to its presentation and/or publication. The opinions or assertions contained herein are the private views of the author, and are not to be construed as official, or as reflecting true views of the Department of the Army or the Department of Defense. The investigators have adhered to the policies for protection of human subjects as prescribed in AR 70–25.

Corresponding Author: Keith R. Lohse, lohse@wustl.edu

Abstract

Contextual interference is one of the most established effects in motor learning research; random practice schedules are associated with poorer performance (in the short-term) but superior learning (in the longer-term) when compared to block practice schedules. However, the way this interference affects learners on a trial-to-trial basis remains poorly understood. We present a secondary data analysis of N=84 healthy young adults, replicating the contextual interference effect in a time estimation task. We used the determinant of a correlation matrix to measure the amount of order in participants' responses. The determinant is conceptually equivalent to the unexplained variance ($1-r^2$) but applies to higher dimensional spaces. We calculated this determinant in two different phase spaces: (1) Trial Space, which was the determinant of the previous 5 trials (lagged constant error 0-4); and (2) Target Space, the determinant of the previous 5 trials of the same target. The distinction in phase space is critical because for blocked practice the previous trial is almost always the same target, but for random practice the previous trial is almost never the same target. In Trial Space, there was no significant difference between groups (p=0.98) and no Group x Lag interaction (p=0.54), although there was an effect of Lag (p<0.01). In Target Space, there were effects of Group (p=0.02), Lag (p<0.01), and a Group x Lag interaction (p=0.03). Participants who practiced using random schedules showed smaller determinants overall, which got smaller as more past trials were included (i.e., increasingly correlated responses). This increase in orderliness was due to the random group having positively correlated errors from trial-to-trial in Target Space. We argue this "response inertia" in the random practice group suggests a greater reliance on the retrieval of the target time from memory. Data from the novel analyses presented herein support the reconstruction account of the contextual interference effect and help integrate the effect with other learning principles in psychology (e.g., retrieval practice being beneficial for long-term recall).

In their seminal study, Shea and Morgan (1979) demonstrated that randomized practice schedules, in which you randomize the order of different tasks, promoted long-term learning at the cost of short-term performance compared to blocked practice conditions. This effect, termed *contextual interference* (CI), explains superior learning as a function of the level of interference that occurs during practice. Random practice schedules create interference because one must switch between different tasks (e.g., ACB-BCA-CAB) during practice, whereas blocked practice leads to less interference because the same task is practiced from trial to trial (e.g., AAA-BBB-CCC). Numerous published reports suggest that the interference produced by random practice schedules during the acquisition phase is beneficial for the long-term retention of motor skills. In contrast, blocked practice has been shown to be beneficial for short-term performance during the acquisition phase (because it produces less interference), but these schedules lead to poorer performance on delayed retention and transfer tests (Merbah & Meulemans, 2011; Broadbent, et al., 2017; Cross et al., 2007).

Although the CI effect is one of the most robust and replicable effects in motor learning, the exact nature of "interference" or precisely why it is beneficial for long-term learning remains unclear (Lee & Simon, 2004; Wymbs & Grafton, 2009). One potentially informative approve to improve understanding of why interference is beneficial would be to study how participants make adjustments from trial to trial during practice. Although it is well documented that random-practice schedules lead to larger errors during practice on average, less research exists exploring how participants respond to and correct errors as a function of their practice schedules. It is possible that random practice is associated with larger errors during practice yet more adaptive corrections from trial to trial. Although there is not much work related to the specific concept of trial-to-trial adjustments as a function of practice schedules, but there is quite a bit of information surrounding it, including research on different types and magnitudes of errors (Lee, et. al., 2016; Albert & Shadmehr, 2016), trial-to-trial adjustments outside of practice scheduling (e.g., van Beers et al., 2015), and how errors during practice relate to exploration of the movement space (e.g., Wu et al., 2014).

Ultimately, the ability to correct errors is a good way of capturing understanding, knowledge, and skill (e.g., Marchal-Crespo, et. al., 2017; Pressing & Rodgers, 1997). 'Errorful' learning can play an especially important role in the consciously mediated stage of learning, such that detecting errors and determining how to correct them are critical components of skill acquisition. Through the experience of error and feedback from the error in a motor task, the motor commands that an individual uses can be updated for the next attempt. If these adjustments are consolidated, this can lead to a more permanent change in the capability for a behavior (i.e., learning; Schmidt, Lee, Winstein, Wulf, & Zelaznik, 2018). Individuals who produce a larger feedback response to error may also be able to learn more than other individuals from a given error (Albert & Shadmehr, 2016). Thus, errors – if successfully detected and adjusted for – are a vital part of the learning process (e.g., Wu et al., 2014; Lohse et al., 2020). Successful movement is about solving motor problems in new situations, not merely engraining the correct (but potentially rigid) movement pattern through repetition (e.g., Bernstein, 1966). As such, exploring how participants respond to errors under different practice schedules may yield important insights into the learning process.

As beneficial as errors may be in learning, there are some instances in which having a very low rate of errors during practice have been shown to be beneficial. In these cases, motor learning tasks are constrained early in practice to minimize performance error while the skill was eventually made more technical. It is commonly hypothesized in these scenarios that with the absence of explicit instruction, minimizing error helps prevent the use of hypothesis testing strategies, which are what ultimately allow participants to correct errors during learning (Maxwell, et al., 2001; Poolton, et al., 2005). In learning environments that minimize error, participants may be able to learn better with smaller errors, because smaller errors were less likely to invoke conscious processing, thereby making participants less likely to engage explicit/declarative approaches to problem solving (Maxwell, et al., 2001).

In the current study, we explored the relationship between practice schedules, adjustments from trial to trial, and long-term learning using an existing dataset. Thomas and colleagues (2021)

demonstrated a contextual interference effect in a time estimation task (see Figure 1). Participants were required to hold a button down for three different target durations, 1500ms, 1700ms, and 1900ms, over 210 practice trials (70 trials at each target). Participants assigned to the blocked schedule performed all trials at a single target before moving onto the next target, with the order of targets counterbalanced across participants. Participants assigned to the random schedule performed all trials in a pseudo-randomized order, with the restriction that targets could not repeat more than once (e.g., AAB, but not AAA). Approximately one day later, participants returned for a delayed retention and transfer test. The retention test consisted of the same targets that participants practiced during acquisition, whereas transfer consisted of two new target times (1600 and 1800 ms). Results from Thomas et al. (2021) replicated the traditional contextual interference effect, with randomly-scheduled practice associated with worse performance during acquisition but superior performance on the retention and transfer-tests (see Figure 1).

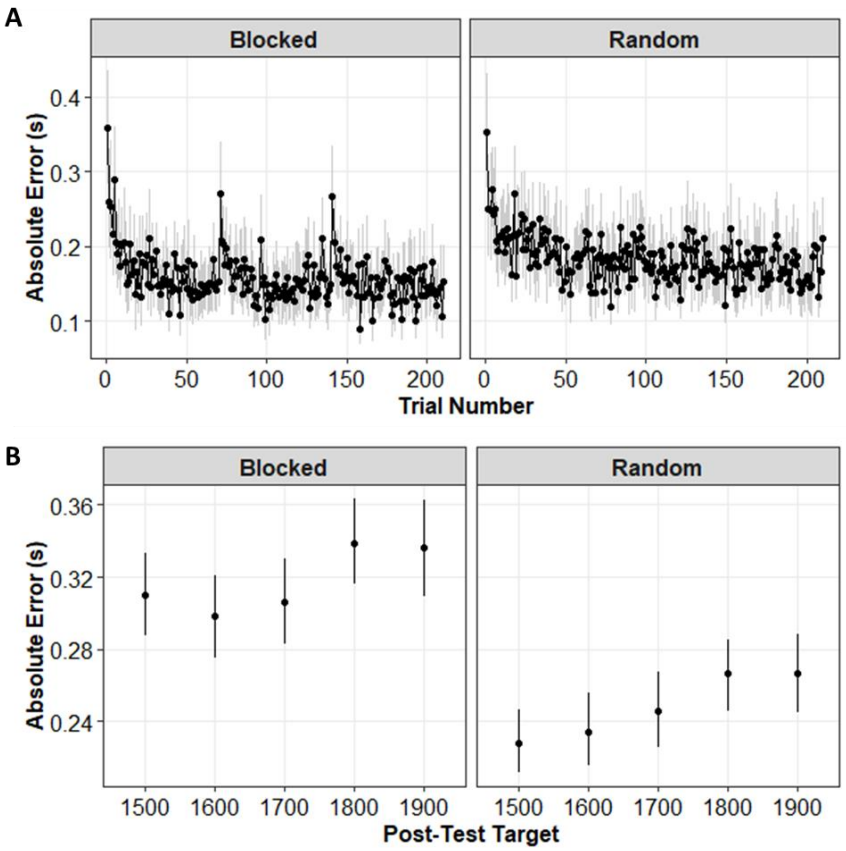

Figure 1. (A) Acquisition data and (B) post-test data from Thomas et al. (2021), showing absolute error as a function practice schedule (blocked versus random) and time in practice (during acquisition) or target (during the post-test). Points show the mean and bars show the 95% confidence interval at each point. Note that 1500, 1700, and 1900ms targets were practiced during acquisition and made up the retention test, 1600 and 1800ms target were not practiced during acquisition and made up the transfer test.

By undertaking a secondary analysis of the data from Thomas et al. (2021), in the present study we explore how differences in practice scheduling affect the way that participants responded to errors. Specifically, although randomly-scheduled participants made larger errors during practice, we hypothesized those participants would be better at correcting those errors. In contrast, we would expect block-scheduled participants to make smaller errors on average but would be worse at correcting those errors. To capture these trial-to-trial corrections, we calculated lagged-variables in two different phase

spaces: Trial Space; and Target Space. In dynamical systems theory, "phase space" refers to a multidimensional space where each dimension represents a degree of freedom of the system. In *Trial Space*, we calculated correlation matrices for the constant error on the current trial (n) and lagged constant error from the previous trial ($n-1$) sequentially back to the fourth previous trial ($n-4$). In *Target Space*, we calculated correlation matrices for the constant error on the current trial (n_k) and lagged constant error for *previous trials of the same target* ($n_k - 1$ to $n_k - 4$). The importance of these two phase-spaces and specific calculations are provided in the Statistical Analysis section, below. In brief, however, we hypothesized that: (1) randomly-scheduled practice would be associated with greater correlations between errors; and (2) following an error, randomly-scheduled participants would make more accurate corrections.

METHODS

Participants

Altogether, 84 healthy young adults (age < 35 years) with no self-reported neurological or musculoskeletal impairments were recruited from the local university population via bulletin posts and word of mouth. Participants were randomly assigned into four training groups differentiated by their training schedule (blocked versus random) and whether they engaged in error estimation during practice or not. The different groups were: (1) blocked with error estimations (M_{age} = 22.62, *SD* = 2.44); (2) blocked without no estimation (M_{age} = 21.43, *SD* = 2.23); (3) random with error estimations (M_{age} = 23.28, *SD* = 4.04); and (4) random no estimation (M_{age} = 21.09, *SD* = 2.53). Although error-estimation was a factor of interest in the primary study (Thomas et al., 2021), there were no statistically significant effects of error estimation in this secondary analysis. Due to this lack of substantial differences, we collapsed across the error estimation factor. Thus, in the results below we consider only two groups, those who had a blocked practice schedule (n=41) and those who had a random practice schedule (n=43).

Five of the 84 participants were primarily left-handed, but all reported their right hand as the preferred hand to control a computer mouse. The experiment was approved by the university's Institutional Review Board (IRB), and written informed consent was obtained from each participant. All participants were naïve to the hypotheses of the experiment. Additionally, the sample size was determined based on past-estimates of contextual interference effects on learning (Brady, 2004), yielding 80% statistical power to show the contextual interference effect in Thomas et al (2021). However, there was no *a priori* power calculation for any of the exploratory analyses.

Task and Stimuli

Details of the task have previously been described in Thomas et al. (2021), so we focus only on the most critical aspects of the methods here. Participants completed a time-estimation task using their dominant hand while seated a computer. The time-estimation task required participants to hold down a mouse button with their index finger for the duration of a target time that was shown on the screen at the beginning of each trial. The target times were 1500, 1700, and 1900 ms. This 200-ms difference was selected based on pilot data, which showed that this subtle distinction was difficult but learnable for the participants (in an effort to avoid floor/ceiling effects).

All participants completed 210 trials during the practice phase, with 70 trials for each target. For participants practicing with a blocked schedule, all 70 trials for the same target were completed together, with the order of the targets counterbalanced across participants. For participants with a random practice schedule, the 70 trials for each target were pseudo-randomly interspersed across the 210 practice trials. This distribution was pseudo-random because targets were constrained that a single target time could not be repeated more than twice in sequence. In both groups, participants received signed error feedback following each trial (e.g., "-125 ms" indicating that a response was slightly too short; "+820 ms" indicating that a response was substantially too long). If participants were within +/-50 ms of the intended target, feedback of "00" was displayed on the screen indicating that the participants were accurate. This 50-ms bandwidth around the target was chosen in an attempt to reduce over-correcting on the part of the

participants (i.e., ±50 ms is likely too small an interval for human nervous system to reliably correct). Following practice, all participants completed the Rating Scales of Mental Effort (RSME; Veltman & Gaillard, 1996), self-reporting their perceived mental effort during practice. Additionally, participants in the error-estimation groups were required to estimate their constant error prior to receiving feedback on every seventh trial. So, for that subset of participants we also have the error-estimate mismatch, defined as the absolute difference between estimated and actual constant error, as a measure of participants awareness of their errors (for more detail, see Thomas et al., 2021).

Approximately 24 hours after practice, participants returned to the laboratory to complete retention and transfer testing. The test consisted of 40 trials, with a set of 20 trials completed in a blocked order and 20 trials completed in a random order. The order of these sets was counterbalanced across participants. In each set, participants completed 4 trials at each of 5 targets: the three original targets (1500, 1700, and 1900 ms) which were considered the retention test and two new targets (1600 and 1800 ms) which are considered the transfer test. Importantly, set order did not have any statistically significant effects in our primary study (Thomas et al., 2021), so we averaged across set order and the individual target times in the present analyses, creating only one experimental factor for the post-tests: retention versus transfer tests.

Trial Phase Space and Target Phase Space during Practice

In order to explore sequential effects during practice, we considered the effect that the practice schedule had on neighboring trials. As shown in Figure 2, there are (at least) two different ways that we can consider the structure of practice. One we will refer to as "Trial Space", where a trial (t_n) is compared to the trial before it (t_{n-1}) or after it (t_{n+1}), regardless of what targets are being practiced on those trials. Alternatively, we can consider these relationships in "Target Space", where a trial of a specific target (t_{n_k}) is compared to the previous trial of the same target (t_{n_k-1}) or the next trial of the same target (t_{n_k+1}), regardless of the absolute trial number.

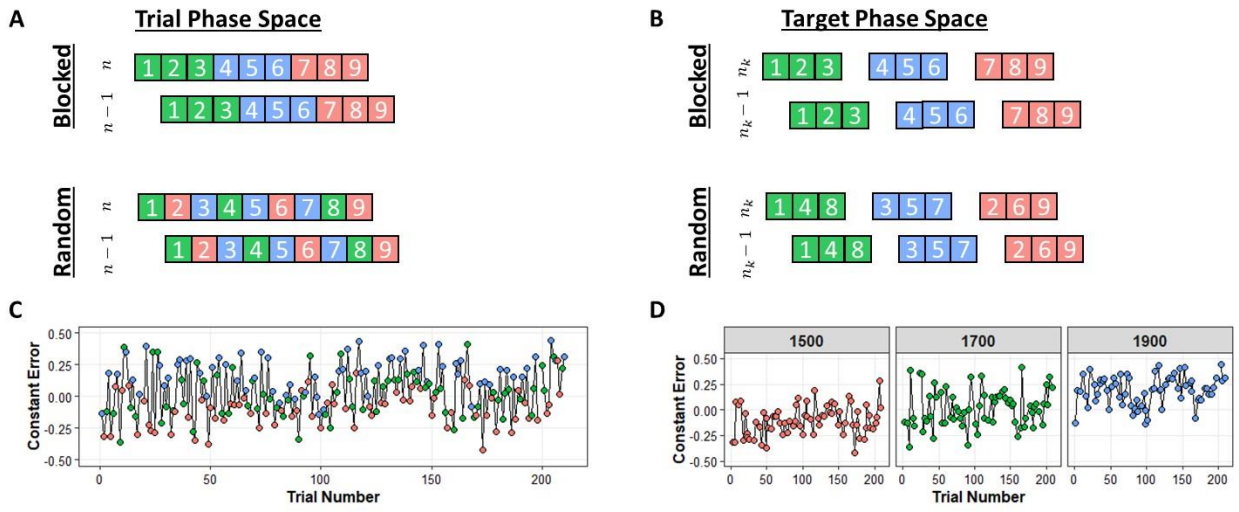

Figure 2. Conceptual diagrams showing the relationship between the current and previous trial in trial phase space (A) and in target phase space (B). Note that when auto-correlations are calculated in trial phase space, $r_{n,n-1}$, the initial trial needs to be dropped from the analysis as there is no previous trial. When the auto-correlation is calculated in target phase space, r_{n_k,n_k-1}, the first trial of each target needs to be dropped as there is no previous trial of that target. The shuffling of the errors is also shown for one randomly scheduled participant's actual data, with constant error across all 210 trials is shown in the original trial phase space (C) and transformed target phase space (D) as a function of target type (red = 1500, green = 1700, and blue =1900 ms).

The distinction between phase spaces is important, because in trial space, the blocked practice group almost never has a trial of one target proceeded or followed by a different target (Figure 2A); this only happens at the boundaries between blocks of trials. In contrast, the random practice group almost never has a trial of one target proceeded or followed by the same target. Indeed, the median number of trials between the same target was 3 and maximum was 9 for the random practice group. These differences mean that when the trials are re-shuffled into target space (Figure 2B), there is very little change in the trial-to-trial relationships for the blocked practice group, but there is a substantial change in the trial-to-trial relationships for the random practice group.

Using both of these phase spaces, we systematically tested whether the relationship between trial-to-trial corrections was different between groups. To capture the correlation between trials, we chose to use the determinant of the constant error (CE) auto-correlation matrix going back five trials in both trial

224 space (CE_n to CE_{n-4}) and target space (CE_{n_k} to CE_{n_k-4}). It is important to first explain why we chose to
225 focus on constant error. Second, it is important to explain why the determinant of the correlation matrix is
226 a useful statistic.

227 First, we chose constant error as our primary outcome because it already takes the target into
228 account, whereas a variable like the hold time on each trial does not (i.e., constant $error_{ij}$ = hold $time_{ij}$ –
229 $target_j$), and because it retains the signed value of the error, whereas a variable like absolute error does not
230 (i.e., absolute $error_{ij}$ = |constant $error_{ij}$|; Schmidt, Lee, Winstein, Wulf, & Zelaznik, 2018). Both of these
231 features are desirable because accounting for the target makes subsequent statistically modeling simpler
232 (i.e., variation due to target is already removed) and retaining the sign makes the correlation between
233 trials more interpretable (i.e., the direction errors, and thus their similarity, cannot be determined from
234 absolute errors alone). Second, we chose the determinant of the constant error correlation matrix because
235 it allows us to capture the structure between errors of multiple, different lags. That is, if we were solely
236 focused on the relationship between the current trial and the previous trial, we could take the correlation
237 coefficient from the lag-1 autocorrelation ($r_{n,n-1}$). However, we wanted to explore the possible
238 relationship between more distant trials, for which we operationally chose a maximum lag of four ($n -$
239 4). Accounting for the relationship between five different trials (i.e., n to $n - 4$), means that our main
240 outcome is not a single correlation, but a correlation matrix. The *determinant* of the correlation matrix
241 thus allows us to reduce any square $n \times n$ matrix into a single scalar value that can be analyzed
242 statistically. As explained below, the determinant is conceptually similar to the unexplained variance,
243 with smaller determinants indicating stronger correlations in the matrix.

244 The relationship of the determinant to unexplained variance is easiest to show in the case of 2×2
245 correlation matrix. The determinant of a 2×2 matrix ($\boldsymbol{A}$) is equal to the product of the diagonal elements
246 minus the product of the off-diagonal elements:

247 (eq1) $$\det(\boldsymbol{A}) = \det\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = ad - bc \ .$$

Thus, in a 2×2 correlation matrix (**R**) the determinant is:

$$\text{(eq2)} \quad \det(\boldsymbol{R}) = \det\left(\begin{bmatrix} 1 & r_{2,1} \\ r_{1,2} & 1 \end{bmatrix}\right) = 1 - r^2$$

making the determinant of a 2×2 correlation matrix mathematically equivalent to the unexplained variance.

As shown in Figure 3, the determinant has a geometric interpretation that we think is useful for generalizing to higher dimensional spaces. Consider the joint-distribution of two uncorrelated normally distributed variables, these uncorrelated data can be captured by a *circle* (e.g., a 95% confidence ellipse is shown in Figure 3A). Next, consider a distribution of two strongly correlated normally distributed variables. These correlated data would be captured by an *ellipse* and the axes of the ellipse are determined by the strength of the correlation (e.g., a 95% confidence ellipse is shown in Figure 3B). The ratio of squared volumes of these two distributions can be shown to equal the determinant of the empirical correlation matrix (Figure 3C). Specific determinants for two different participants (one with a blocked schedule and one with a random schedule) are shown in Figure 3D-E. In 3D, constant error is plotted as a time series for each participant. In 3E, the lag-1 autocorrelation is shown in target space, $r(n_k, n_k\text{-}1)$, for each participant. The participant who had a blocked schedule showed almost no correlation between current and previous error, making the explained variance very small, $r^2 < 0.01$, and thus the determinant very large, $d > 0.99$. In contrast, the participant who had a random schedule showed a modest correlation between current and previous error, yielding an $r^2 = 0.15$, and thus the determinant $d = 0.85$.

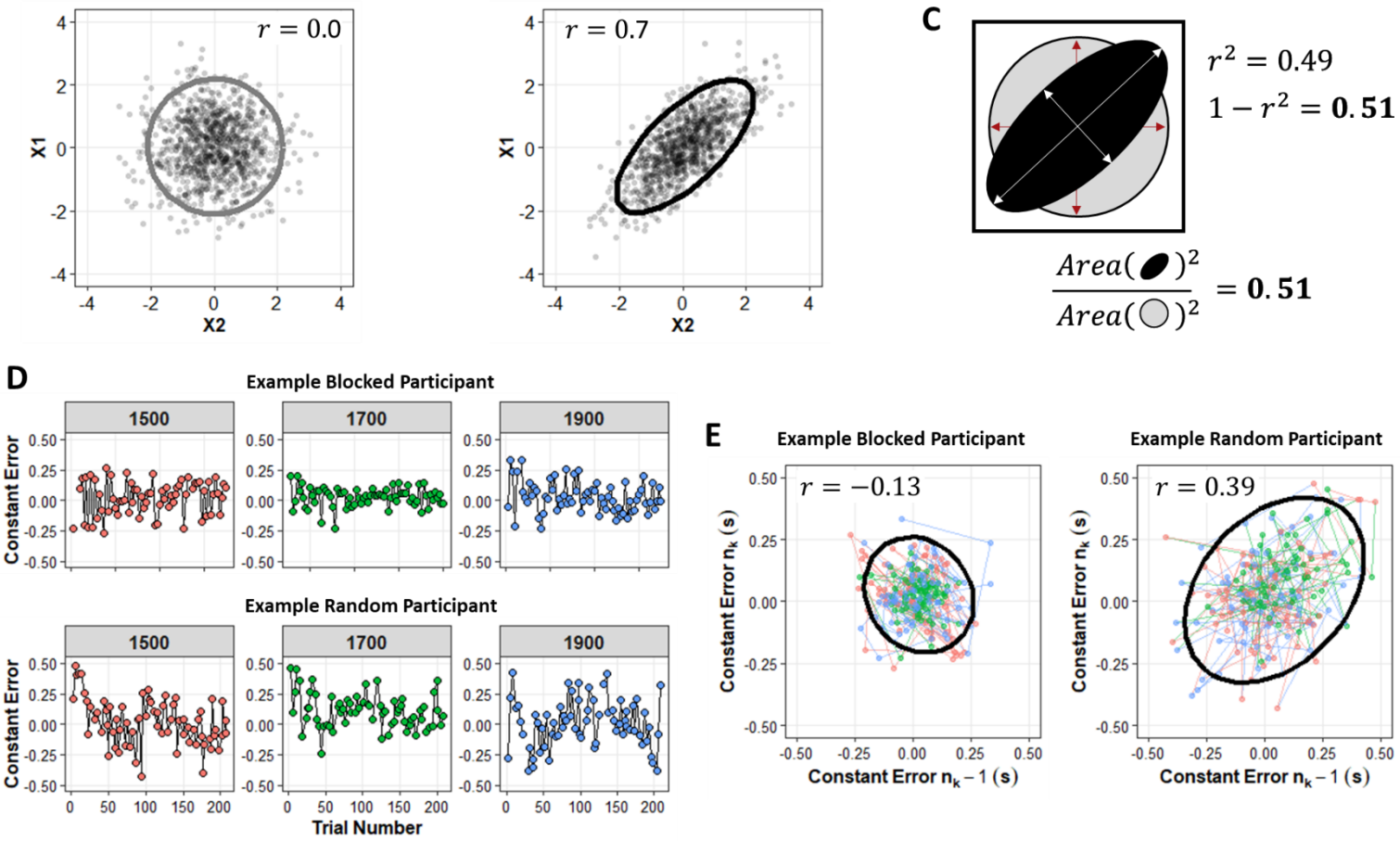

Figure 3. The geometric interpretation of the determinant for a 2 × 2 correlation matrix. (A) The circular 95% confidence region for n=1,000 uncorrelated data points. (B) The elliptical 95% confidence region for n=1,000 correlated data points where r=0.7. (C) The ratio of the squared area of these regions (0.51) is equivalent to the determinant of the correlation matrix, [1 0.7; 07 1], which is 0.51. For reference, arrows show the major and minor axes of the circle (red) and ellipse (white). (D) Example time series for one block-schedule participant and one random-schedule participant. (E) Scatter plots showing the lag-1 autocorrelation for the same block- and random-schedule participants with a 95% confidence ellipse and the Pearson's *r* value calculated in target space.

In sum, the determinant tells us how the volume of a unit square is transformed by a given matrix (Margalit & Rabinoff, 2017). When applied to a correlation matrix, the determinant can tell us how much this volume shrinks based on the strength of the correlation (see also Lohse, Jones, Healy, & Sherwood, 2014). Although this is typically shown with squares and parallelograms in linear algebra, it also holds for circles and ellipses when applied to normally distributed random variables (i.e., the major and minor axes are being transformed in a similar way). In two dimensions, the determinant reflects an *ellipse* whose area is dictated by the strength of a correlation ($r_{1,2}$) relative to a *circle* (the alternative distribution which assumes $r_{1,2} = 0$). In three dimensions, the determinant would reflect an *ellipsoid* whose volume is dictated by all three correlations ($r_{1,2}, r_{1,3}, r_{2,3}$) relative to a *sphere* (the alternative distribution which assumes all $r's = 0$). With more than three dimensions, the geometric interpretation is difficult (nigh impossible) to visualize, but the interpretation still holds: the determinant reflects the ratio of the volume taken up by the observed distribution relative to what it would be if the variables were all independent. Thus, the determinant is bounded between 0 and 1, with a smaller determinant meaning that more variance has been explained.

Statistical Analysis

All data processing, analysis, and visualization were done in R 4.0.4 and R Studio (RStudio Team, 2020; Wickham et al., 2019). Code and de-identified data for these analyses are available from: https://github.com/keithlohse/taylor_2022_CI_sequential_effects. To analyze the correlations between errors, we calculated determinants using different numbers of lagged trials from one trial back ($n - 1$) to four trials back ($n - 4$), in both trial space and target space for each subject. These determinants were then analyzed using a mixed-factorial repeated measures ANOVA with a between-participants factor of Group (blocked versus random practice schedules) and within-participants factors of Phase Space (target versus trial) and Lag (including 1, 2, 3, or 4 of the previous trials in the correlation matrix). Mauchly's test was used to assess violations of sphericity, and the Greenhouse-Geisser correction was applied when sphericity was violated (denoted by p_{gg}, Lawrence, 2016).

To determine the way in which prior trials related to future trials, we followed this analysis of determinants with mixed-effect regressions (Bates, Maechler, Bolker, &Walker, 2015), where the constant error on the current trial was regressed onto constant errors from the previous trial(s). Full details of the mixed-effect regression models are presented in Supplemental Appendix i, but in brief, constant error on the next trial was regressed onto fixed-effects of constant error from the previous four trials. The model also included a random intercept for each participant and random slopes of for each lagged constant error variable. These random-effects account for the within-participant nature of the data (Long, 2012; Snijders & Bosker, 2011). As detailed in the Results, the only significant differences between groups were in the immediately previous trial (lag-1 error), so we focused on only the immediately previous trial in subsequent analyses.

Regressing constant error onto past constant errors tells us how similar past errors are to each other. That is, a positive slope would indicate that positive errors are followed by positive errors and negative errors by negative errors. However, we also have to account for the fact that errors move around the zero-point, making the absolute distance from zero on the next trial meaningful. That is, moving from -150 ms to -50 ms is arguably as "good" of a correction as moving from -150ms to +50 ms. In order to address that issue, we also regressed *absolute* error on the next trial onto constant error from the previous trial. Because absolute error showed a u-shaped curvilinear relationship with previous constant error (i.e., large negative or positive constant errors were followed by large absolute errors), we also included a quadratic fixed-effect of previous constant error in the model. As before, random-effects included a random intercept and slopes for the linear and quadratic effects of previous constant error to account for the within-subject nature of these data. Statistical significance of these effects was determined using the Welch-Satterthwaite approximation to the degrees of freedom (Kuznetsova, Brockhoff, Christensen, 2017). To ensure robustness of results, we used semi-parametric bootstrapping to estimate 95% confidence intervals for all model parameters (Bates, Maechler, Bolker, &Walker, 2015).

Finally, we present some exploratory regression results demonstrating how individual differences in the determinant during practice relate to individual differences in long-term retention and transfer, self-reported mental effort, and error estimation accuracy (for those participants who were forced to estimate their errors). For these analyses, we regressed different dependent variables onto the determinant in target space and Group (Random versus Blocked). As with the mixed-effect regressions, full details of these models are presented in Supplemental Appendix i.

RESULTS

Correlations between Trials during Practice

The Group x Phase Space x Lag mixed-factorial ANOVA for the determinant of the correlation matrix yielded several statistically significant effects: a main-effect of Lag, $F(3,246)=165.03$, $p_{gg}<0.001$, a main-effect of Phase Space, $F(1,82)=10.15$, $p_{gg}=0.002$, a Group x Phase Space interaction, $F(1,82)=7.04$, $p_{gg}=0.010$, a Lag x Phase Space interaction, $F(3,246)=9.83$, $p_{gg}=0.002$, and most critically a Group x Lag x Space interaction, $F(3,246)=4.39$, $p_{gg}=0.036$.

To unpack this three-way interaction, we ran post-hoc Group x Lag mixed-factorial ANOVAs in trial space and target space separately. As shown in Figure 4, in trial space there was a non-significant effect of Group, $F(1,82)=<0.01$, $p=0.981$, a significant main effect of Lag, $F(3,246)=152.31$, $p_{gg}<0.001$, and a non-significant Group x Lag interaction, $F(3,246)=0.40$, $p_{gg}=0.541$. Thus, in trial space, there was greater order in responses when more previous trials were included, but this increase in order did not significantly differ as a function of practice schedule. In target space, however, there was a significant main effect of Group, $F(1,82)=5.09$, $p_{gg}=0.027$, a main-effect of Lag, $F(3,246)=120.17$, $p_{gg}<0.001$, and a Group x Lag interaction, $F(3,246)=4.29$, $p_{gg}=0.039$. Thus, in target space, although both groups tended to have increasingly correlated responses when more previous trails were considered, this effect was stronger for the random practice group.

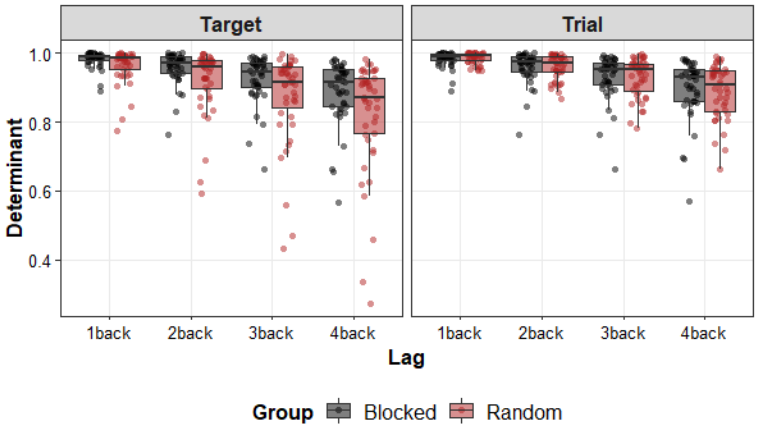

Figure 4. The determinants of the correlation matrix as a function of Group, Phase Space, and Lag (the number of previous trials included in the correlation matrix).

Exploring the Nature of Adjustments from Trial-to-Trial

Correlation Matrices. Although the determinant reflects the amount of unexplained variance in a correlation matrix, it does not tell us the specific directions or magnitudes of the correlations involved. Thus, although we know that the random-practice schedule was associated with more correlated errors from trial-to-trial (i.e., more order in participants' responses), it does not tell us specifically *how* an error on the previous trial relates to an error on the next trial. To understand the trial-to-trial adjustments better, we present three different analyses. First, as shown in Table 1, we present the average correlations between trials as a function of practice schedule and phase space as descriptive statistics. Although all of these correlations tend to be small (*r's* < 0.20), the largest correlations were found for the random practice group in target space where weak positive correlations were common (*r*'s between 0.10 and 0.15) and generally double to triple the correlations found in other groups/phase spaces.

Table 1. The correlation matrices for constant error in the five previous trials as a function of phase space and group.

Random Group in Target Space							**Random Group in Trial Space**					
	Nk	**Nk-1**	**Nk-2**	**Nk-3**	**Nk-4**			**N**	**N-1**	**N-2**	**N-3**	**N-4**
Nk	1	0.137	0.108	0.094	0.096		**N**	1	0.041	0.079	0.057	0.055
Nk-1	0.137	1	0.137	0.106	0.093		**N-1**	0.041	1	0.045	0.081	0.055
Nk-2	0.108	0.137	1	0.140	0.105		**N-2**	0.079	0.045	1	0.047	0.083

Nk-3	0.094	0.106	0.140	1	0.136		**N-3**	0.057	0.081	0.047	1	0.048
Nk-4	0.096	0.093	0.105	0.136	1		**N-4**	0.055	0.055	0.083	0.048	1
Blocked Group in Target Space							**Blocked Group in Trial Space**					
	Nk	**Nk-1**	**Nk-2**	**Nk-3**	**Nk-4**			**N**	**N-1**	**N-2**	**N-3**	**N-4**
Nk	1	0.043	0.052	0.050	0.051		**N**	1	0.040	0.053	0.050	0.045
Nk-1	0.043	1	0.051	0.050	0.060		**N-1**	0.040	1	0.047	0.050	0.056
Nk-2	0.052	0.051	1	0.052	0.061		**N-2**	0.053	0.047	1	0.047	0.059
Nk-3	0.050	0.050	0.052	1	0.056		**N-3**	0.050	0.050	0.047	1	0.053
Nk-4	0.051	0.060	0.061	0.056	1		**N-4**	0.045	0.056	0.059	0.053	1

*****Shaded regions denote correlation coefficients $r>0.10$. All cells show the Pearson correlation coefficient on average across participants.

Constant Error on the Next Trial. Mixed-effect regressions predicting constant error on the next trial from constant error on the previous four trials showed differential effects in trial space relative to target space. (Full details of the regression models are available in Supplemental Appendix i.) In trial space, there were statistically significant main-effects of Group ($p<0.001$), Lag-1 error ($p=0.002$), Lag-2 error ($p<0.001$), Lag-3 error ($p<0.001$), and Lag-4 error ($p<0.001$). Critically however, there were no Group x Lag interactions for either Lag-1 error ($p=0.953$), Lag-2 error ($p=0.250$), Lag-3 error ($p=0.637$), or Lag-4 error ($p=0.917$). These results can be seen in the dashed lines of Figure 5A; random practice participants generally had more positive constant errors than blocked practice participants, but the effect of the previous trial was comparable across groups (only Lag-1 error is shown).

In target space, there were statistically significant main-effects of Group ($p<0.001$), Lag-1 error ($p<0.001$), Lag-2 error ($p<0.001$), Lag-3 error ($p<0.001$), and Lag-4 error ($p<0.001$). Critically there was also a statistically significant Group x Lag-1 error interaction ($p=0.005$), but no other Group x Lag interactions, Lag-2 error ($p=0.244$), Lag-3 error ($p=0.628$), or Lag-4 error ($p=0.204$). These results can be seen in the solid lines of Figure 5A; random practice participants not only had more positive constant errors than blocked practice participants, but random practice participants also tended to have more similar errors from one trial to the next compared to blocked practice participants (note the more positive slope of the solid line for the random group compared to the blocked group).

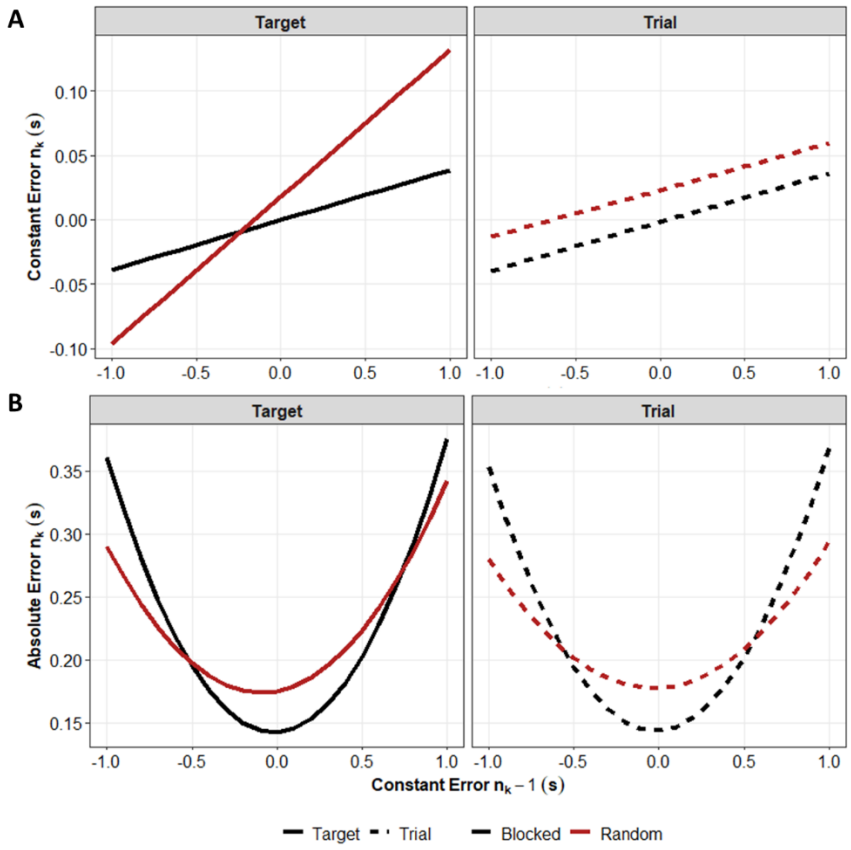

Figure 5. The model predictions for constant error on the next trial (A) or absolute error on the next trial (B) as a function of the previous constant error. Coefficients for all of the models are provided in the supplemental appendix. Solid lines indicate predictions from the model in target space, dashed lines indicate model predictions in trial space. Red lines show model predictions for the random practice group, Black lines show model predictions for the blocked practice group.

Absolute Error on the Next Trial. Mixed-effect regressions predicting absolute error on the next trial from constant error on the previous trial showed slightly different effects in trial space relative to target space. In trial space, there was a statistically significant main-effect of Group (p<0.001), no linear effect of Lag-1 error (p=0.221), and a significant quadratic effect of Lag-1 error (p<0.001). Although there was not a significant Group x Lag-1 interaction (p=0.967), there was a significant interaction with the quadratic effect, Group x Lag-1^2 (p<0.001). Participants who practiced with a random schedule tended to make larger errors on the subsequent trial and, although both groups showed u-shaped distributions to their corrections, the u-shape for the blocked practice participants was tighter and deeper than the u-shape

for the random practice participants; see Figure 5B. For reference, about 95% of the errors fell between -500 ms and +500 ms, so the group difference is especially crucial in that range.

In target space, there was a statistically significant main-effect of Group (p=0.003), linear Lag-1 error (p=0.004), and quadratic Lag-1^2 error (p<0.001). Although there was not a significant Group x Lag-1 interaction (p=0.103), there was a significant interaction with the quadratic effect, Group x Lag-1^2 (p=0.025). As shown in Figure 5B, participants who practiced with a random schedule tended to make larger errors on the subsequent trial and, although both groups showed u-shaped distributions to their corrections, the u-shape for the blocked practice participants was tighter and deeper than the u-shape for the random practice participants. Interestingly, compared to trial space, there was evidence for a "tilt" in these distributions (shown by the linear effect of Lag-1 error) such that both groups tended to make slightly larger absolute errors following positive constant errors compared to negative constant errors.

Associations (or lack thereof) with Long Term Learning

Retention Test. A multivariable regression model in which average absolute error on the retention test was regressed onto Group and the Determinant over the previous 5 trials in target space showed that there was a statistically significant main-effect of Group, b=-0.08, t(1,81)=-2.71, p=0.008, but not a statistically significant main-effect of the Determinant, b=-0.04, t(1,81)=-0.43, p=0.672. Collinearity for these predictors was relatively low, with variance inflation factor = 1.06. A scatterplot illustrating these effects is shown in Figure 6A.

Transfer Test. A multivariable regression model in which average absolute error on the transfer test was regressed onto Group and the Determinant over the previous 5 trials in target space demonstrated that there was a statistically significant main-effect of Group, b=-0.07, t(1,81)=-2.66, p=0.009, but not a statistically significant main-effect of the Determinant, b=-0.01, t(1,81)=-0.13, p=0.896. A scatterplot illustrating these effects is shown in Figure 6B.

Self-Reported Mental Effort. Average mental effort as self-reported on the Rating Scales of Mental Effort was regressed onto Group and the Determinant over the previous 5 trials in target space showed that there was not a statistically significant main-effect of Group, b=-7.42, t(1,81)=-1.43, p=0.156, and a marginally significant effect of the Determinant, b=-37.99, t(1,81)=-2.02, p=0.047. However, given the large p-value and a lack of predictions for this association, we did not interpret this effect further. A scatterplot illustrating these effects is shown in Figure 6C.

Error Estimation Accuracy. For participants who estimated their own errors (N=42), we similarly regressed error estimation accuracy onto Group and the Determinant over the previous 5 trials. There was no statistically significant main-effect of Group, b=21.08, t(1,39)=1.05, p=0.299, and no statistically significant main-effect of the Determinant, b=-48.56, t(1,38)=-0.72, p=0.476. A scatterplot illustrating these effects is shown in Figure 6D.

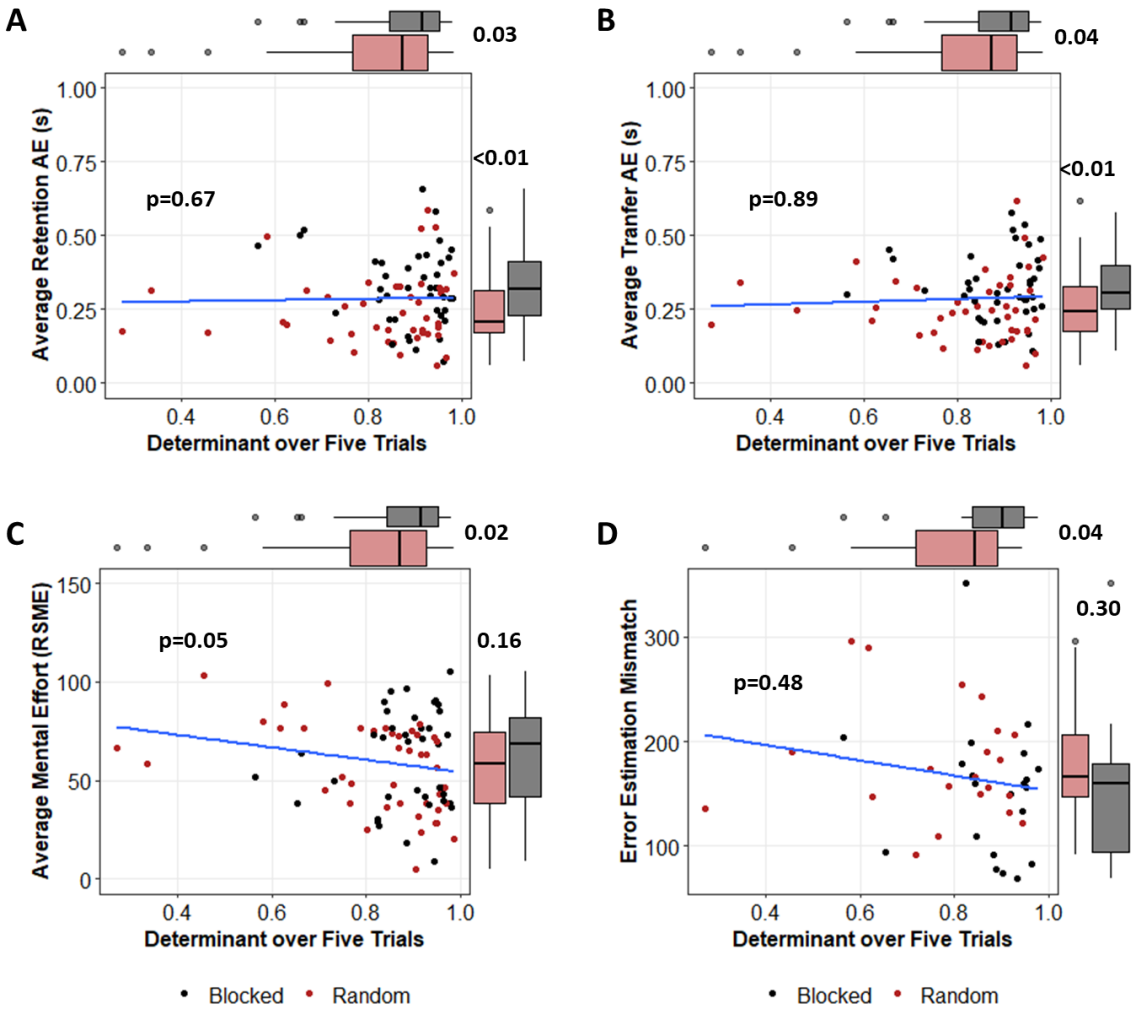

Figure 6. The average absolute error (AE) during retention (A) and transfer tests (B), plus the average from the rating scales of mental effort (RMSE; C), and the mis-match between actual error and estimated error (D) as a function of the determinant in target space and group. *P*-values are given in the margins for the effect of Group controlling for the other variable (i.e., the difference in retention test performance had p=0.03 controlling for the determinant; the difference in the determinant had p<0.01 controlling for retention test performance). The *p*-value in the plot is given for the association between the variable of interest (A-D) and the determinant, controlling for Group.

DISCUSSION

In this study, we report that random practice schedules are associated with greater order in responses (i.e., stronger correlations as shown by the determinant) in target space than in trial space. In contrast, the blocked practice group showed very little difference in correlations between trial space and target space. For random practice participants, these correlations were quite small and positive (*r*'s between 0.10 to 0.15), but notably larger than the correlations in either phase space or for blocked practice participants (*r*'s between 0.00 to 0.05).

These findings for the determinants of the correlation matrix supported our first hypothesis that a random practice schedule would be associated with stronger correlations (i.e., more orderly/systematic responding) from trial to trial. However, we did not find support for our second hypothesis that random practice schedules would be associated with more adaptive corrections from trial to trial. In contrast, random practice was associated with positive correlations between errors, such that if a participant overshot on the previous trial, they were more likely to overshoot on the next trial (as shown in Figure 5A). Moreover, although random practice participants did tend to reduce their error from trial to trial, participants with a blocked schedule were better at making adaptive corrections (i.e., a smaller absolute error on trial n+1 given the same constant error on trial n, see Figure 5B).

Thus, several interesting patterns emerge when we consider the sequential constant error and sequential absolute error effect together: (1) random practice schedules do lead to adaptive corrections (i.e., absolute error is more likely to be smaller on the next trial), but the type of error will be similar to error that came before (i.e., positive correlations between constant errors); (2) blocked practice schedules lead to *more* adaptive corrections (i.e., even smaller absolute errors on the subsequent trial), but the nature of the previous error as little to do with the nature of the subsequent error (i.e., null-correlations between constant errors); and (3) trial-to-trial corrections for the blocked practice participants, in either phase space, resembled corrections for the random practice participants in trial space, not target space.

The third suggests that in trial space, random practice participants have little use for the error from the previous trial to inform their response on the next trial, because that trial is of a different target. In target space, in contrast, that error is actually useful for updating the internal representation of the target time to improve performance the next time that target is seen. For blocked participants, however, regardless of the space the subsequent trial is (almost) always the same target as the previous trial. Why then do blocked participants behave like random participants in trial space (when internal updating has no benefit between trials) rather than random participants in target space (when internal updating has a practical benefit between trials)?

We speculate that blocked practice leads participants to respond more to the feedback itself rather than to use that feedback to update an internal representation of the target time. This finding is most consistent with the *forgetting-reconstruction hypothesis* of the CI effect, which states that a previously constructed action plan is more likely to be available in working memory during blocked practice, whereas in random practice, the individual is forced to forget the action plan because they must move on to a different trial, thus, needing to reconstruct the action plan the next time around (Lee & Magill, 1983; 1985). That is, randomly scheduled participants appear to be using both the memory of their last response (reflected in positive correlations), plus the feedback they received (reflected in reduced absolute error), in order to make their correction on the next trial. In contrast, block scheduled participants appear to be only using the feedback to guide their response. This creates a sort of "response inertia" in the random practice participants, who move closer to the target time, but are slow to adapt; in other words, overshoots are likely followed by smaller overshoots, undershoots by smaller undershoots.

The finding that slower adapters show better long-term retention has been demonstrated in other motor learning and adaptation tasks (Smith et al., 2006; Coltman, Cashaback & Gribble, 2019). Motor learning is not a singular process, with many computational models suggesting that adaptation is the result of multiple learning processes each with their own, distinct timescales (Smith et al., 2006; Lee and Schweighofer, 2009; Haith & Krakauer, 2013). For instance, trial-to-trial variation in motor adaptation

tasks is well characterized by a model with two processes that each have a "retention" parameter (how much learning is preserved from one trial to the next) and a "learning rate" parameter (how much a learner changes the movement in response to an error). The "fast" learning process learns quickly but has low retention whereas the slow process learns slowly by has higher retention. Some researchers have posited that this "slow" learning process is responsible for chronic changes in behavior over longer periods (e.g., improvement in average performance from Day 1 to Day 2), whereas the "fast" learning process is responsible for acute changes in behavior (e.g., faster acquisition or "savings" in practice on Day 2 compared to Day 1; Albert & Shadmehr, 2018; McDougle et al., 2015), although some data suggest the slow process contributes to both (Coltman et al., 2019).

These multi-process learning models have been applied to contextual interference effects before (Schweighofer, Lee, Goh, et al., 2011; Kim, Oh & Schweighofer, 2015). Schweighofer, Lee, Goh, et al. (2011) replicated the traditional contextual interference effect in able-bodied adults and in a sample of adults with stroke (>3 months post-stroke). In the sample of adults with stroke, individual differences in visuospatial working memory modulated long-term learning with a blocked schedule, but not a random schedule. Specifically, in the blocked practice group, individuals with worse working memory actually showed better retention, whereas individual differences in working memory did not explain retention following randomly scheduled practice. This paradoxical result was accounted for by a computational model that contained a fast process and multiple slow processes. In an "unimpaired" model where the fast process was intact, the fast process learns quickly to improve performance, however, this reduces the error-driven updating of the slow processes and thus led to worse long-term retention. When a visuospatial working memory deficit is simulated by "impairing" the fast process, this leads to more persistent errors, giving the slow process the information it needs to adapt and improve retention.

Although we did not employ a multi-process computational model in our analysis, the results of our statistical models provide conceptually similar results while also yielding some complementary new insights. Specifically, we our data reinforce that being slow to adjust performance is associated with

improved long-term learning at a group-level. (Although our regressions did not find evidence that individual differences in the determinant related to individual differences in learning, as discussed in the limitations below.) Our analyses extend this past-work, however, by showing the different relationships between consecutive errors in both trail space and target space, whereas past work (including computational models) have focused on trial space (e.g., Kim, Oh & Schweighofer, 2015; Pauwels, Swinnen & Beets, 2014). This phase space difference for the random practice group suggests that the response to errors is not simply governed by passive memory processes with different timescales, but active psychological processes in which errors from a particular target are encoded and retrieved the next time they see a stimulus of the same target (Lee & Magill, 1983; 1985).

Limitations

Although our novel secondary analysis provides some potential insights into the contextual interference effect, it is important to emphasize that these findings are primarily "hypothesis generating" in nature and need to be confirmed in independent samples (see Tukey, 1980; Wagenmakers et al., 2012). Similarly, although the primary study was powered to detect a contextual interference effect defined as the difference between blocked- and random-practice groups on the delayed retention/transfer tests (Thomas et al., 2021), there was no *a priori* power calculation for the myriad statistical tests we conducted in this secondary analysis. As such, statistically significant results (like the difference in determinants between groups during practice) need to be replicated and non-significant results need to be treated with caution. For instance, at the group level, random practice was associated with better long-term retention and transfer, and with greater correlations between sequential errors during practice in the short term. However, in our regression analyses, there was not a statistically significant relationship between individual differences in the determinant and individual differences in learning after controlling for practice group, as shown in Figure 6A/B. Given the absence of an informed power analysis, we cannot say whether this lack of statistically significant effects is due to a lack of statistical power or to a genuine lack of an effect. Similarly, we face a major validity issue if we think about the determinant of the

correlation matrix in target space as “the” way to capture interference captured by practice scheduling. Although we saw group-level differences in learning and the determinant, part of the reason we saw no significant associations between learning and the determinant at the individual-level may be that the determinant is not the best way to operationalize the construct that we are really interested in. That is, the determinant tells us how errors are correlated during practice, but may not be the best way to capture how participants are actually perceiving errors and/or making updates to any sort internal model.

Conclusions

In conclusion, we found that randomly scheduled practice was associated with stronger correlations between errors during practice, but we did not find evidence that random practice was associated with better corrections from trial-to-trial. Thus, practicing with a random-schedule led to errors on the next trial that were generally smaller but similar to errors on the previous trial, whereas practice with a blocked schedule led to much smaller errors on the next trial that were not reliably correlated with the error from the previous trial. This “response inertia” on the part of randomly scheduled participants is consistent with the forgetting and reconstruction account of the contextual interference effect.

REFERENCES

Albert, S. T., & Shadmehr, R. (2016). The neural feedback response to error as a teaching signal for the motor learning system. *Journal of Neuroscience*, *36*(17), 4832-4845.

Albert, S. T., & Shadmehr, R. (2018). Estimating properties of the fast and slow adaptive processes during sensorimotor adaptation. *Journal of Neurophysiology*, *119*(4), 1367-1393. https://doi.org/10.1152/jn.00197.2017

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1-48. doi:10.18637/jss.v067.i01.

Bernstein, N. (1966). The co-ordination and regulation of movements. *The co-ordination and regulation of movements*. Pergamon Press, London.

Brady, F. (2004). Contextual interference: a meta-analytic study. *Perceptual and motor skills*, *99*(1), 116-126.

Broadbent, D. P., Causer, J., Williams, A. M., & Ford, P. R. (2017). The role of error processing in the contextual interference effect during the training of perceptual-cognitive skills. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(7), 1329-1342. http://dx.doi.org/10.1037/xhp0000375

Coltman, S. K., Cashaback, J. G., & Gribble, P. L. (2019). Both fast and slow learning processes contribute to savings following sensorimotor adaptation. *Journal of Neurophysiology*, *121*(4), 1575-1583. https://doi.org/10.1152/jn.00794.2018

Cross, E. S., Schmitt, P. J., & Grafton, S. T. (2007). Neural substrates of contextual interference during motor learning support a model of active preparation. *Journal of Cognitive Neuroscience*, *19*(11), 1854–1871.

Galea, J. M., Vazquez, A., Pasricha, N., Orban de Xivry, J. J., & Celnik, P. (2011). Dissociating the roles of the cerebellum and motor cortex during adaptive learning: The motor cortex retains what the cerebellum learns. *Cerebral Cortex*, *21*(8), 1761-1770.

Haith, A. M., & Krakauer, J. W. (2013). Model-based and model-free mechanisms of human motor learning. In *Progress in motor control* (pp. 1-21). Springer, New York, NY.

Kim, S., Oh, Y., & Schweighofer, N. (2015). Between-trial forgetting due to interference and time in motor adaptation. *PloS one*, *10*(11), e0142963. https://doi.org/10.1371/journal.pone.0142963

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest package: tests in linear mixed effects models. *Journal of Statistical Software*, *82*, 1-26.

Lawrence, M. A., & Lawrence, M. M. A. (2016). Package 'ez'. *R package version*, *4*(0).

Lee, T. D., Eliasz, K. L., Gonzalez, D., Alguire, K., Ding, K., & Dhaliwal, C. (2016). On the role of error in motor learning. *Journal of Motor Behavior*, *48*(2), 99-115. DOI: 10.1080/00222895.2015.1046545

Lee, T. D., & Magill, R. A. (1983). The locus of contextual interference in motor-skill acquisition. *Journal of Experimental Psychology: Human Learning and Memory*, *9*(4), 730–746. https://doi.org/10.1037/0278-7393.9.4.730

Lee, T. D., & Magill, R. A. (1985). Can forgetting facilitate skill acquisition? In D. Goodman, R. B. Wilberg, & I. M. Franks (Eds.), *Differing perspectives in motor learning, memory, and control* (pp. 3–21). Elsevier Science.

Lee, J. Y., & Schweighofer, N. (2009). Dual adaptation supports a parallel architecture of motor memory. *Journal of Neuroscience*, *29*(33), 10396-10404. https://doi.org/10.1523/JNEUROSCI.1294-09.2009

Lee, T. D., & Simon, D. A. (2004). Contextual interference. In N. J. Hodges & A. M. Williams (Eds.), *Skill acquisition in sport: Research, theory and practice, (pp. 29-44)*. Routledge.

Lohse, K. R., Jones, M., Healy, A. F., & Sherwood, D. E. (2014). The role of attention in motor control. *Journal of Experimental Psychology: General*, *143*(2), 930. https://doi.org/10.1037/a0032817

Lohse, K.R., Miller, M.W., Daou, M., Valerius, W., Jones, M. (2020). Dissociating the contributions of reward-prediction errors to trial-level adaptation and long-term learning. *Biological Psychology*, *149*, 107775. DOI: 10.1016/j.biopsycho.2019.107775

Long, J. D. (2012). *Longitudinal data analysis for the behavioral sciences using R*. Sage.

Marchal-Crespo, L., Michels, L., Jaeger, L., López-Olóriz, J., & Riener, R. (2017). Effect of error augmentation on brain activation and motor learning of a complex locomotor task. *Frontiers in Neuroscience*, *11*, 526. doi: 10.3389/fnins.2017.00526

Margalit, D., Rabinoff, J., & Rolen, L. (2017). Interactive linear algebra. *Georgia Institute of Technology*.

Maxwell, J. P., Masters, R. S. W., Kerr, E., & Weedon, E. (2001). The implicit benefit of learning without errors. *The Quarterly Journal of Experimental Psychology Section A*, *54*(4), 1049-1068.

McDougle, S. D., Bond, K. M., & Taylor, J. A. (2015). Explicit and implicit processes constitute the fast and slow processes of sensorimotor learning. *Journal of Neuroscience*, *35*(26), 9568-9579. doi: 10.1523/JNEUROSCI.5061-14.2015

Merbah, S., & Meulemans, T. (2011). Learning a motor skill: Effects of blocked versus random

practice a review. *Psychologica Belgica*, *51*, 15.

Pauwels, L., Swinnen, S. P., & Beets, I. A. (2014). Contextual interference in complex bimanual skill learning leads to better skill persistence. *PloS one*, *9*(6), e100906. https://doi.org/10.1371/journal.pone.0100906

Poolton, J. M., Masters, R. S. W., & Maxwell, J. P. (2005). The relationship between initial errorless learning conditions and subsequent performance. *Human Movement Science*, *24*(3), 362-378.

Pressing, J., & Jolley-Rogers, G. (1997). Spectral properties of human cognition and skill. *Biological Cybernetics*, *76*(5), 339-347.

RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/.

Schmidt, R. A., Lee, T. D., Winstein, C., Wulf, G., & Zelaznik, H. N. (2018). *Motor control and learning: A behavioral emphasis*. Human Kinetics.

Shea, J. B. & Morgan R. L. (1979). Contextual interference effects on the acquisition, retention, and transfer of a motor skill. *Journal of Experimental Psychology: Human Learning and Memory*, *5*(2), 179-187.

Schweighofer, N., Lee, J. Y., Goh, H. T., Choi, Y., Kim, S. S., Stewart, J. C., ... & Winstein, C. J. (2011). Mechanisms of the contextual interference effect in individuals poststroke. *Journal of Neurophysiology*, *106*(5), 2632-2641. DOI: 10.1152/jn.00399.2011

Smith, M. A., Ghazizadeh, A., & Shadmehr, R. (2006). Interacting adaptive processes with different timescales underlie short-term motor learning. *PLoS Biology*, 4(6), e179.

Snijders, T. A., & Bosker, R. J. (2011). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. sage.

Thomas, J.L., Fawver, B., Taylor, S., Miller, M.W., Williams, A.M., & Lohse, K.R., (2021). Using error-estimation to probe the psychological processes underlying contextual interference effects. *Human Movement Science*, *79*, 102854. https://doi.org/10.1016/j.humov.2021.102854

Tukey, J. W. (1980). We need both exploratory and confirmatory. *The American Statistician*, *34*(1), 23-25.

van Beers, R. J., van der Meer, Y., & Veerman, R. M. (2013). What autocorrelation tells us about motor variability: insights from dart throwing. *PloS ONE*, 8(5), e64332. https://doi.org/10.1371/journal.pone.0064332

Wagenmakers, E. J., Wetzels, R., Borsboom, D., van der Maas, H. L., & Kievit, R. A. (2012). An agenda for purely confirmatory research. *Perspectives on Psychological Science*, *7*(6), 632-638. https://doi.org/10.1177/1745691612463078

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D. A., François, R., ... & Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), 1686.

Wu, H. G., Miyamoto, Y. R., Castro, L. N. G., Ölveczky, B. P., & Smith, M. A. (2014). Temporal structure of motor variability is dynamically regulated and predicts motor learning ability. *Nature Neuroscience*, *17*(2), 312-321.

Wymbs, N. F., & Grafton, S. T. (2009). Neural substrates of practice structure that support future off-line learning. *Journal of Neurophysiology*, *102*(4), 2462-2476. https://doi.org/10.1152/jn.00315.2009

Supplemental Appendix i

Supplemental Table i. The details of the mixed-effect model regressing constant error (on trial n) to constant error on the previous trials (n-1 to n-4) in TARGET space.

```
          AIC      BIC   logLik deviance df.resid
 -4183.8  -3976.4   2118.9  -4237.8    16025

Scaled residuals:
    Min      1Q  Median      3Q     Max
-4.7626 -0.5828 -0.0166  0.5431  4.8947

Random effects:
 Groups      Name                        Variance  Std.Dev. Corr
 participant (Intercept)                 5.444e-04 0.023332
             target_lag_constant_error   8.112e-03 0.090069  0.08
             target_lag_2_constant_error 6.328e-04 0.025155  0.17 -0.06
             target_lag_3_constant_error 1.666e-03 0.040820 -0.17  0.51  0.32
             target_lag_4_constant_error 2.016e-03 0.044905  0.03  0.34  0.64  0.75
 block       (Intercept)                 5.808e-05 0.007621
 Residual                                4.429e-02 0.210458
Number of obs: 16052, groups:  participant, 84; block, 3

Fixed effects:
                                          Estimate Std. Error         df t value Pr(>|t|)
(Intercept)                             -1.863e-04  6.207e-03  8.457e+00  -0.030 0.976756
groupRandom                              1.801e-02  6.146e-03  7.784e+01   2.930 0.004453 **
target_lag_constant_error                3.866e-02  1.929e-02  1.012e+02   2.004 0.047756 *
target_lag_2_constant_error              4.544e-02  1.279e-02  9.195e+01   3.552 0.000606 ***
target_lag_3_constant_error              5.263e-02  1.377e-02  9.439e+01   3.821 0.000238 ***
target_lag_4_constant_error              4.226e-02  1.392e-02  9.415e+01   3.037 0.003094 **
groupRandom:target_lag_constant_error    7.505e-02  2.609e-02  9.255e+01   2.876 0.004992 **
groupRandom:target_lag_2_constant_error  1.989e-02  1.695e-02  8.047e+01   1.173 0.244297
groupRandom:target_lag_3_constant_error  8.921e-03  1.835e-02  8.378e+01   0.486 0.628121
groupRandom:target_lag_4_constant_error  2.379e-02  1.857e-02  8.352e+01   1.281 0.203644
```

Supplemental Table ii. The details of the mixed-effect model regressing constant error (on trial n) to constant error on the previous trials (n-1 to n-4) in TRIAL space.

```
     AIC      BIC   logLik deviance df.resid
 -3637.9  -3429.4   1846.0  -3691.9    16667

Scaled residuals:
    Min      1Q  Median      3Q     Max
-4.7347 -0.5817 -0.0178  0.5390  5.2021

Random effects:
 Groups      Name                        Variance  Std.Dev. Corr
 participant (Intercept)                 0.0007328 0.027070
             trial_lag_constant_error    0.0054379 0.073742  0.00
             trial_lag_2_constant_error  0.0016769 0.040950  0.11  0.18
             trial_lag_3_constant_error  0.0031155 0.055816 -0.03  0.18 -0.03
             trial_lag_4_constant_error  0.0029850 0.054635 -0.08  0.36 -0.13  0.13
 block       (Intercept)                 0.0000438 0.006618
 Residual                                0.0461369 0.214795
Number of obs: 16694, groups:  participant, 84; block, 3

Fixed effects:
                                          Estimate Std. Error         df t value Pr(>|t|)
(Intercept)                              -0.001718   0.006193  12.762623  -0.277 0.785948
groupRandom                               0.024816   0.006840  78.668848   3.628 0.000506 ***
trial_lag_constant_error                  0.037837   0.017204 100.844536   2.199 0.030139 *
trial_lag_2_constant_error                0.048082   0.013833  95.761160   3.476 0.000767 ***
trial_lag_3_constant_error                0.048786   0.015180 106.682187   3.214 0.001733 **
trial_lag_4_constant_error                0.038497   0.014926 108.943604   2.579 0.011236 *
groupRandom:trial_lag_constant_error     -0.001376   0.023121  90.504484  -0.060 0.952682
groupRandom:trial_lag_2_constant_error    0.021270   0.018350  83.326439   1.159 0.249731
groupRandom:trial_lag_3_constant_error   -0.009601   0.020281  94.759363  -0.473 0.637021
groupRandom:trial_lag_4_constant_error    0.002094   0.019955  96.726839   0.105 0.916654
```

Supplemental Table iii. The details of the mixed-effect model regressing absolute error (on trial n) to constant error on the previous trial (n-1) in TARGET space.

```
        AIC      BIC   logLik deviance df.resid
-18821.1 -18704.9   9425.6 -18851.1    17094

Scaled residuals:
    Min      1Q  Median      3Q     Max
-2.9510 -0.6639 -0.1948  0.4438  6.3231

Random effects:
 Groups      Name                          Variance  Std.Dev. Corr
 participant (Intercept)                   2.148e-03 0.046351
             target_lag_constant_error     5.141e-04 0.022674  0.09
             I(target_lag_constant_error^2) 1.294e-02 0.113737 -0.42  0.05
 block       (Intercept)                   7.303e-05 0.008546
 Target      (Intercept)                   8.354e-05 0.009140
 Residual                                  1.906e-02 0.138067
Number of obs: 17109, groups:  participant, 84; block, 3; Target, 3

Fixed effects:
                                          Estimate Std. Error        df t value Pr(>|t|)
(Intercept)                               0.142482   0.010370 17.911193  13.740 5.95e-11 ***
groupRandom                               0.032299   0.010407 83.227325   3.103  0.00261 **
target_lag_constant_error                 0.007441   0.008447 88.774258   0.881  0.38077
I(target_lag_constant_error^2)            0.225397   0.027528 73.655317   8.188 5.89e-12 ***
groupRandom:target_lag_constant_error     0.018580   0.011272 80.058541   1.648  0.10319
groupRandom:I(target_lag_constant_error^2) -0.084151   0.036798 68.725145  -2.287  0.02529 *
```

Supplemental Table iv. The details of the mixed-effect model regressing absolute error (on trial n) to constant error on the previous trial (n-1) in TRIAL space.

```
   AIC      BIC   logLik deviance df.resid
-18621.3 -18504.9   9325.6 -18651.3    17256

Scaled residuals:
    Min      1Q  Median      3Q     Max
-2.5028 -0.6633 -0.1972  0.4395  6.2349

Random effects:
 Groups      Name                          Variance  Std.Dev. Corr
 participant (Intercept)                   2.093e-03 0.045748
             trial_lag_constant_error      8.279e-04 0.028774  0.15
             I(trial_lag_constant_error^2) 1.022e-02 0.101081 -0.30 -0.15
 block       (Intercept)                   8.121e-05 0.009011
 Target      (Intercept)                   9.494e-05 0.009744
 Residual                                  1.949e-02 0.139596
Number of obs: 17271, groups:  participant, 84; block, 3; Target, 3

Fixed effects:
                                           Estimate Std. Error         df t value Pr(>|t|)
(Intercept)                               0.1437246  0.0106181 15.9366985  13.536 3.72e-10 ***
groupRandom                               0.0337704  0.0102837 83.1466501   3.284  0.00150 **
trial_lag_constant_error                  0.0076990  0.0090065 91.7551952   0.855  0.39487
I(trial_lag_constant_error^2)             0.2171921  0.0262886 71.9049775   8.262 5.02e-12 ***
groupRandom:trial_lag_constant_error     -0.0005079  0.0120604 83.0754443  -0.042  0.96651
groupRandom:I(trial_lag_constant_error^2) -0.1076113  0.0350303 67.1134214  -3.072  0.00307 **
```

Supplemental Table v. The details of the ordinary least-squares regression model predicting average absolute error at retention as a function of practice schedule and the determinant in target space.

```
lm(formula = ave_ae_Retention ~ rand.c + det_Target.c, data = MERGED)

Residuals:
     Min       1Q   Median       3Q      Max
-0.24519 -0.09090 -0.02349  0.07920  0.33953

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.28631    0.01382  20.711  < 2e-16 ***
rand.c       -0.07722    0.02845  -2.714  0.00811 **
det_Target.c -0.04392    0.10331  -0.425  0.67188
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Supplemental Table vi. The details of the ordinary least-squares regression model predicting average absolute error at transfer as a function of practice schedule and the determinant in target space.

```
lm(formula = ave_ae_Transfer ~ rand.c + det_Target.c, data = MERGED)

Residuals:
     Min       1Q   Median       3Q      Max
-0.20994 -0.07518 -0.01438  0.07885  0.36314

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.28693    0.01251  22.928  < 2e-16 ***
rand.c       -0.06859    0.02575  -2.663  0.00934 **
det_Target.c -0.01227    0.09352  -0.131  0.89597
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Supplemental Table vii. The details of the ordinary least-squares regression model predicting average error estimation mismatch as a function of practice schedule and the determinant in target space.

```
lm(formula = EEM ~ rand.c + det_Target.c, data = MERGED)

Residuals:
    Min      1Q  Median      3Q     Max
-89.571 -43.007  -2.689  32.887 196.624

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   164.228      9.579  17.144   <2e-16 ***
rand.c         21.080     20.033   1.052    0.299
det_Target.c  -48.560     67.438  -0.720    0.476
```

Supplemental Table viii. The details of the ordinary least-squares regression model predicting average ratings of mental effort as a function of practice schedule and the determinant in target space.

```
lm(formula = ME_AVE ~ rand.c + det_Target.c, data = MERGED)

Residuals:
    Min      1Q  Median      3Q     Max
-50.003 -18.078   0.626  18.760  47.240

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)    58.755      2.519  23.321   <2e-16 ***
rand.c         -7.421      5.185  -1.431   0.1562
det_Target.c  -37.990     18.828  -2.018   0.0469 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```