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 cite the pre-print 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: July 11th, 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. The authors would like to thank two anonymous reviewers for their constructive and thoughtful critiques. We would also like to thank Dr. Tim Lee for comments he provided on an earlier version of this pre-print.

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

Abstract

Contextual interference is an established phenomenon in learning research; random practice schedules are associated with poorer performance, but superior learning compared to blocked practice schedules. We present a secondary 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 participant responses. We calculated this determinant in different phase spaces: Trial Space, the determinant of the previous 5 trials (lagged constant error 0-4); and Target Space, the determinant of the previous 5 trials of 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). Ultimately, randomly scheduled practice was associated with adaptive corrections but positive correlations between errors from trial to trial (e.g., overshoots followed by smaller overshoots). Blocked practice was associated with more adaptive corrections but uncorrelated responses. Our findings suggest that random practice leads to the retrieval and updating of the target from memory, facilitating long term retention and transfer.

In their seminal study, Shea and Morgan (1979) demonstrated that randomized practice schedules, in which you change the order of different tasks from trial to trial, promoted long-term learning at the cost of short-term performance when 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). There are dominant explanations for the effect. First, the *elaboration* hypotheses (Shea & Zimny, 1983; 1988), which broadly argues that switching between different tasks (or different parameters of the same task) with a random schedule makes the difference between tasks more pronounced/salient to the learner. Second, the *reconstruction* hypotheses (Lee & Magill, 1983; 1985), which broadly argues that actively forgetting and then retrieving a motor program (or variations of that program) facilitates later recall. These hypotheses are not mutually exclusive, as demonstrated by Li and Wright (2000) who showed that random-practice schedules interfered with the performance of a secondary task both prior to response initiation (when the motor program is theoretically being retrieved from memory) and during the inter-stimulus interval (when prior actions can theoretically be contrasted against each other).

One potentially informative approach to understanding why interference is beneficial for learning is to study how participants adjust performance 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 but more *adaptive* corrections from trial to trial. For instance, the average magnitude of errors with random schedules is larger, but participants could move closer to the target on the next trial. With blocked schedules, in contrast, the average magnitude of errors may be smaller, but participants could randomly bounce around the target from trial to trial (similar to *maladaptive* corrections as defined in Schmidt, Young, Swinnen, & Shapiro, 1989).

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). Moreover, the relationship between errors during practice and long-term learning has a detailed and complex history. Thorndike (1927) emphasized the role of "correct" feedback to reinforce the preceding motor response and argued that repetition of the correct movement was essential for consolidation into long-term memory (see also Adams, 1968). An alternative view places a greater focus on errors themselves. Bernstein, (1966) argued that successful movement is about solving motor problems in new situations, not merely engraining the correct (but potentially rigid) movement pattern through repetition. This viewpoint underscores the need to recall information from memory and is supported by work showing that providing less feedback can actually be beneficial for learning (e.g., see Lee & Carnahan, 1990; Winstein & Schmidt, 1990) and that recall practice is generally beneficial for long-term retention (in verbal and cognitive learning more generally; e.g., see Bjork, 1988; Roediger & Butler, 2011). These views need not be exclusive, however, as successes and errors can both provide valuable signals for updating internal representations that are retrieved from – and then encoded back into – long-term memory (e.g., reinforcement- and supervised-learning mechanisms working together under the umbrella of motor learning; Haith & Krakauer, 2013; Lohse, Miller, Bacelar, & Krigolson, 2019).

Past-work on the frequency of feedback is very informative in this regard. In naturalistic settings, learners will often get some feedback during or after every motor attempt (as intrinsic vision, proprioception, etc., are available to detect errors). In the laboratory, however, researchers can manipulate the presence and relative frequency of feedback. Published reports show that withholding knowledge of results (KR[1]) can have a beneficial effect for learning (e.g., Winstein & Schmidt, 1990). Lee and Carnahan (1990) manipulated the frequency of feedback by providing bandwidth KR. If participants were inside the margin of error on a trial, no KR was provided (implying success); if participants were outside the margin of error, then KR was given as the signed magnitude of the error. Thus, the wider the bandwidth (margin of error), the less KR participants received during practice. Lee and Carnahan (1990) yoked half of their participants in each bandwidth condition to the feedback schedule of another participant in that condition, dissociating the bandwidth effect from the relative frequency of feedback. Learning with bandwidth KR led to more accurate and stable performance, above and beyond the reduced frequency effect. The authors also demonstrated a novel method for capturing the *adaptive* behavior of their participants following KR and no-KR trials by measuring the absolute change in participants responses from one trial to the next. Following "correct" no-KR trials, participants should attempt to reproduce the same response, yielding a mean change close to 0. Following incorrect trials with KR, participants should change their response, yielding a mean change >0 (and ideally moving closer to the target). An exploratory analysis presented in their discussion section precisely showed this adaptive behavior; participants made smaller absolute changes following no-KR trials compared to trials with KR.

In the current study, we explored the relationship between practice schedules, adjustments from

[1] Knowledge of results (KR) is defined as information about the outcome of a movement and contrasted against knowledge of performance (KP) which is defined information about the quality of the movement. E.g., in dart throwing, KR would be the final landing place of the dart; KP would be the mechanics of the throw itself.

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). Thomas et al. (2021) replicated the traditional contextual interference effect, with randomly-scheduled practice leading to worse performance during acquisition but superior performance on the retention and transfer-tests (see Figure 1).

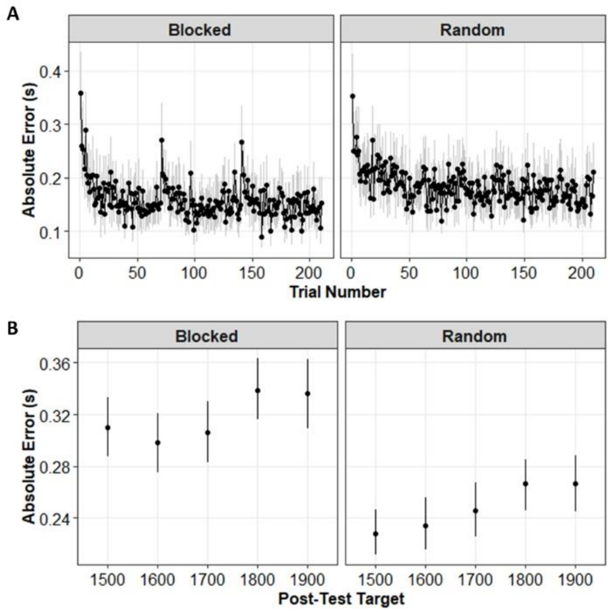

Figure 1. (A) Acquisition data and (B) post-test data from Thomas et al. (2021), showing absolute error as a function practice schedule (blocked versus random) and time in practice (during acquisition) or target (during the post-test). Points show the mean and bars show the 95% confidence interval at 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.

In the present study, we explore how differences in practice scheduling affect the way that participants respond to errors using a secondary analysis of the data reported by Thomas et al. (2021). To capture these trial-to-trial corrections, we calculated lagged-variables in two different phase spaces: *trial* space; and *target* space. Borrowing a term from 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.

Similar to Lee and Carnahan (1990), we also calculated the absolute change in performance from trial to trial, but we specifically did so in target space, $|CE_{n_k} - CE_{n_k-1}|$.[2] Note that Lee and Carnahan used a KR manipulation with a single task practiced across trials, so the trial space/target space distinction is irrelevant in that study. For the present study, however, it is important to make this distinction. We are interested in the response change the next time participants see a stimulus of the same target, thus target space is more important than trial space for these analyses. Thomas et al. (2021) included an error bandwidth of +/50 ms around each target, allowing us to see how participants changed their responses on "correct" no-KR trials and on incorrect trials in which signed KR feedback was given. On no-KR trials, an adaptive response would be making no change, yielding an absolute change ~0. On trials with KR, an adaptive response would be to make a change proportional to the previous error, yielding a slope ~1 when absolute change is regressed onto previous absolute error. (For instance, a |150| ms error on trial $n-1$ should be changed by ~|150| ms on trial n).[3]

Using these autocorrelations and absolute changes in performance from trial to trial, we can test important hypotheses related to reconstruction (Lee & Magill, 1983; 1985) and elaboration explanations of the CI effect (Shea & Zimny, 1983; 1988). From a *reconstruction* perspective, we would predict positive autocorrelations between successive trials for random-practice participants because an internal representation of the task is retrieved from memory, potentially modified based on feedback, and then updated prior to the next trial. Assuming that this updating is neither perfect (e.g., +150 ms does not get corrected by -150 ms precisely) nor overly aggressive (e.g., +150 ms does not get corrected by -500 ms), successive errors should be positively correlated as participants gradually update their internal

---

[2] Absolute performance change was calculated as constant error on the current trial minus constant error on the last trial of the same target, $|CE_{n_k} - CE_{n_k-1}|$. Note that this is mathematically equivalent to simply taking the difference between overall response times, R, after the target, T, is taken into account:

$$\begin{aligned} CE_{n_k} - CE_{n_k-1} &= (R_{n_k} - T_k) - (R_{n_k-1} - T_k) \\ &= (R_{n_k} - R_{n_k-1}) + (T_k - T_k) \\ &= (R_{n_k} - R_{n_k-1}) + (0) \end{aligned}$$

[3] These hypotheses could also hypothetically be tested using signed values rather than absolute values, but previous absolute error makes the analysis simpler (e.g., previous constant error would require nonlinear models to account for -/+ errors). For simplicity, we thus focus on absolute error in these exploratory analyses.

representation of the target (e.g., +150 ms is followed by +120 ms). From an *elaboration* perspective, we would expect more adaptive corrections for random-practice participants, particularly late in practice when internal representations of the task have theoretically been "sharpened" by the random practice schedule. For instance, a random practice participant might still make larger errors overall, but changes in their responses should be more proportional than changes for blocked participants. This effect follows because if participants have better internal representations of the target times (e.g., 'what is 1,500 ms?') then they should be better equipped to make specific responses to feedback (e.g., 'what is 150 ms?').

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). 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 presented here.

Task and Stimuli

The task is described in Thomas et al. (2021), so we present only on the most critical aspects of the methods. Participants completed a time-estimation task using their dominant hand while seated at 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, thereby reducing the risk of floor/ceiling effects.

All participants completed 210 trials during the practice phase, with 3 Sets of 70 trials. 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 such 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 to reduce “maladaptive corrections” (Schmidt, Young, Swinnen, & Shapiro, 1989, p. 358) on the part of the participants (e.g., <50 ms is too small an interval for human nervous system to reliably correct).

Approximately 24 hours after practice, participants returned to the laboratory to complete retention and transfer testing. The test consisted of 40 trials with no KR, 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, namely, retention versus transfer tests.

Trial Phase Space and Target Phase Space during Practice

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. Trial space where a trial (n) is compared to the trial before it ($n - 1$) or after it ($n + 1$), regardless of what targets are being practiced on those trials; and target space, where a trial of a specific target (n_k) is compared to the previous trial of the same target ($n_k - 1$) or the next trial of the same target ($n_k + 1$), regardless of the absolute trial number.

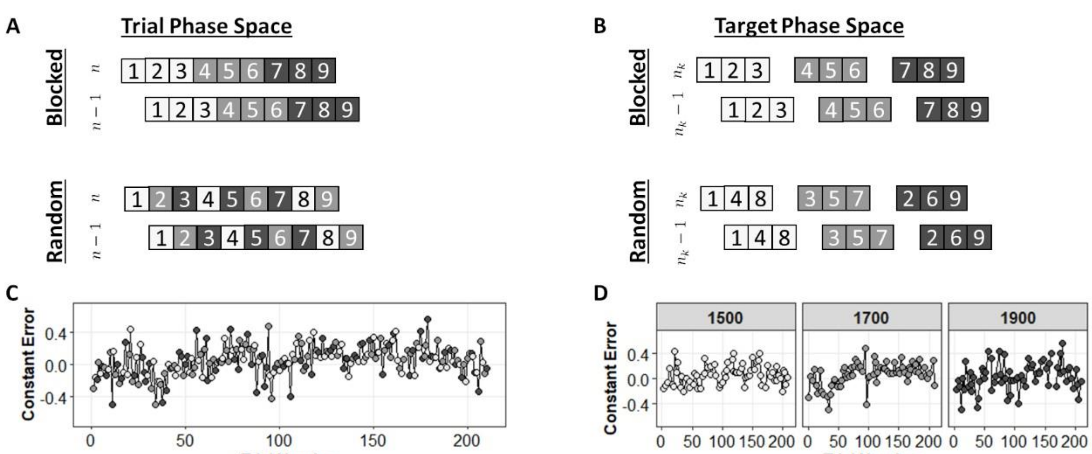

Figure 2. A representation of the conceptual relationship between the current and previous trial in trial phase space (A) and in target phase space (B). Note that when auto-correlations are calculated in trial phase space, $r_{n,n-1}$, the initial trial needs to be dropped from the analysis as there 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 (C) the original trial space and (D) transformed target space as a function of target type (light fill = 1500, medium = 1700, and dark =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. 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 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 four trials in both trial 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 focus on constant error. Second, it is important to explain why the determinant of the correlation matrix is a useful statistic.

First, we chose constant error as our primary outcome because it already takes the target into account, whereas a variable like the response time on each trial does not (i.e., $CE_{nk} = R_{nk} - T_k$); and because it retains the signed value of the error, whereas a variable like absolute error (AE) does not (i.e., $AE_{nk} = | R_{nk} - T_k |$).[4] These features are desirable because accounting for the target makes subsequent

[4] Note that we are referring to CE and AE for a single trial, hence the "n" subscript. This is slightly different from influential definitions given in Schmidt & Lee (2011) where "CE" and "AE" are actually average measurements aggregated across trials (see p. 30). Schmidt and Lee also describe an aggregate measure called "absolute constant error", which they denote ACE or $|CE|$, based on their formulation of CE. Again, however, this is an aggregate measure and distinct from the single trial CE, AE, and absolute change measures in the current study.

statistical modeling simpler (i.e., variation due to target is already removed) and retaining the sign makes the correlation between trials more interpretable (i.e., the direction errors, and thus their similarity, cannot be determined from absolute errors alone). Second, we chose the determinant of the constant error correlation matrix because it allows us to capture the structure between errors of multiple, different lags. That is, if we were solely focused on the relationship between the current trial and the previous trial, we could take the correlation coefficient from the lag-1 autocorrelation ($r_{n,n-1}$). However, we wanted to explore the possible relationship between more distant trials, for which we operationally chose a maximum lag of four ($n-4$). Accounting for the relationship between five different trials (i.e., n to $n-4$), means that our main outcome is not a single correlation, but a correlation matrix. The *determinant* of the correlation matrix thus allows us to reduce any square $n \times n$ matrix into a single scalar value that can be analyzed statistically. As explained below, the determinant is conceptually similar to the unexplained variance, with smaller determinants indicating stronger correlations in the matrix.

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

(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:

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

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 $d = 0.85$.

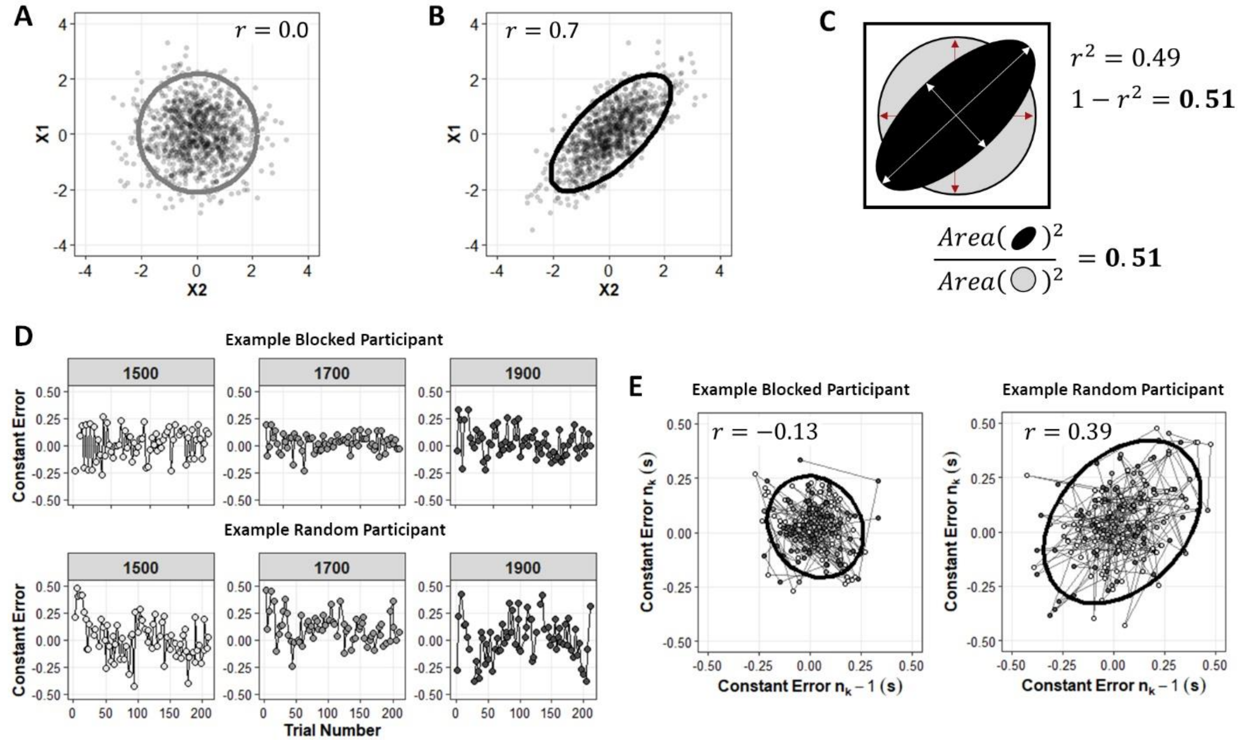

Figure 3. The geometric interpretation of the determinant for a 2 × 2 correlation matrix. (A) The circular 95% confidence region for n=1,000 uncorrelated data points. (B) The elliptical 95% confidence region for n=1,000 correlated data points where r=0.7. (C) The ratio of the squared area of these regions (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. Lines in the scatterplot show "paths" connecting successive trials.

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. 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 to four trials back, in both trial space and target space for each participant. 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-participant 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 how participants adapted their performance based on previous errors, we conducted a series of mixed-effect regressions (Bates, Maechler, Bolker, &Walker, 2015). In the first model, the goal was to analyze how participants changed their performance following KR versus no-KR trials. We aggregated data to obtain the mean change following KR and the mean change following no-KR for each participant. The mean absolute change in performance was then regressed onto factors of Group (Random versus Blocked practice), Set of trials (1, 2, or 3), whether or not KR was present on the previous trial (KR versus no KR), and the interactions of these factors. Random-intercepts were included to account for the within-subject nature of the Set and KR factors (full details are presented in Supplemental Appendix i).

In the second model, we excluded no-KR trials to focus only on those trials when participants received feedback about their error. The absolute change in performance on each trial was regressed onto absolute error from the previous trial, termed Lag AE. Inspecting this relationship within each participant showed that the best fitting model included linear (Lag AE), quadratic (Lag AE^2), and cubic (Lag AE^3) terms. Polynomial effects of Lag AE were then included with fixed effects of Group, Set, and all Group × Set × Lag AE interactions. Random-intercepts and slopes were included to account for the within-participant nature of the Set factor and inter-participant variability in the Lag AE effects (see full details in Supplemental Appendix i). Statistical significance of these effects was determined using the Welch-Satterthwaite approximation to the degrees of freedom (Kuznetsova, Brockhoff, Christensen, 2017).[5]

Finally, we conducted exploratory regressions to see how the determinant of the correlation matrix and the intercepts/slopes from our mixed-effect regressions related to long-term learning. We regressed the average absolute error from the retention tests onto either: (1) the determinant of the correlation matrix in trial space, which tells how errors were correlated from trial to trial; (2) the mean change following trials without KR, which tells us how stable participants' responses were following

[5] Note that additional exploratory analyses were included in our original pre-print (https://doi.org/10.51224/SRXIV.143). We have also included these analyses in Supplemental Appendix ii for transparency about the total number and type of statistical tests conducted.

"correct" feedback; and (3) the estimated slope from the mixed-model, which tells us how proportionally a participant would change their performance given their previous error. All models also controlled for the between subject factor of group. Statistical significance across all models was set to $\alpha = 0.05$ using ANOVA with Type III sums of squares (Fox & Weisberg, 2018).

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 main effects for Lag, $F(1.0, 85.5)=165.03$, $p_{gg}<0.001$, Phase Space, $F(1,82)=10.15$, $p=0.002$, and interactions for Group x Phase Space, $F(1,82)=7.04$, $p_{gg}=0.010$, Lag x Phase Space, $F(1.1, 88.4)=9.83$, $p_{gg}=0.002$, and Group x Lag x Space, $F(1.1, 88.4)=4.39$, $p_{gg}=0.036$.

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 for Group, $F(1,82)=<0.01$, $p=0.981$, a significant main effect for Lag, $F(1.1, 87.7)=152.31$, $p_{gg}<0.001$, and a non-significant Group x Lag interaction, $F(1.1, 87.7)=0.40$, $p_{gg}=0.541$. 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 for Group, $F(1,82)=5.09$, $p=0.027$, a main-effect for Lag, $F(1.0, 85.3)=120.17$, $p_{gg}<0.001$, and a Group x Lag interaction, $F(1.0, 85.3)=4.29$, $p_{gg}=0.039$. Thus, in target space, although both groups tended to have increasingly correlated responses when more previous trials were considered, this effect was stronger for the random practice group.

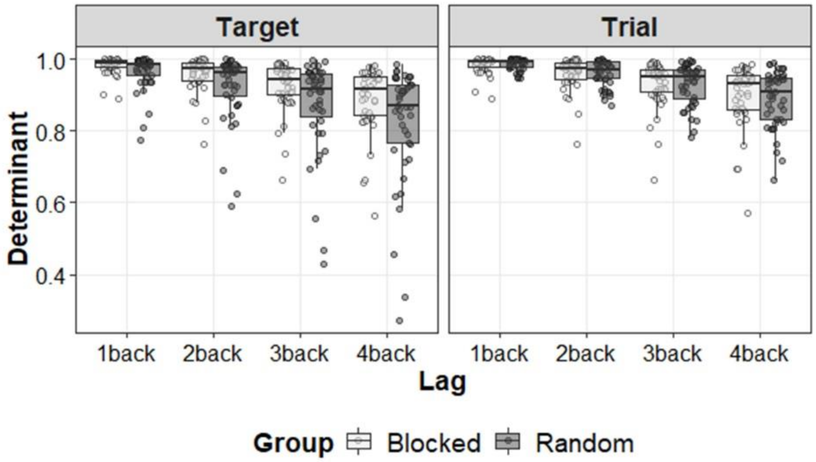

Figure 4. The determinants of the correlation matrix as a function of Group, Phase Space, and Lag (the number of previous trials included in the correlation matrix).

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, 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. These correlations tended to be small (*r's* < 0.20), but the largest correlations were found for the random practice group in target space (*r*'s between 0.10 and 0.15) and were 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 average Pearson correlation coefficient across participants.

Changes in Performance following Errors

Changes following KR versus No KR. Participants were within the 50-ms target bandwidth on 20.8% of trials (following exclusion of outliers) and therefore received no-KR on those trials. Thus, there were also 79.2% of trials on which participants did receive KR (following exclusions). By group, participants with a blocked schedule had 24% “correct” trials with no KR and 76% trials with KR; participants with a random schedule had 18% “correct” trials with no KR and 82% trials with KR.

As shown Figure 5A, participants tended to respond to KR in an adaptive way, making smaller adjustments following “correct” no-KR trials and larger adjustments following incorrect trial when they received KR. Our first mixed-effects regression model yielded statistically significant main-effects for KR, $F(1,82)=356.5$, $p<0.001$, and Set, $F(2,164)=10.07$, $p<0.001$. However, these effects were further superseded by a significant Set × KR interaction, $F(2,164)=9.22$, $p<0.001$, such that the difference between KR and no-KR trials got smaller from Set 1 to Set 2 ($p=0.086$) and from Set 1 to Set 3 ($p=0.027$). This difference across sets was because changes tended to get smaller following KR trials (means = 258, 229, 217 ms), whereas changes following no KR trials stayed relatively constant (means = 158, 156, 153 ms).

Additionally, there was a statistically significant main-effect for Group, $F(1,82)=4.98$, $p=0.028$, but no statistically significant interactions with Group (p’s>0.54), such that participants with a blocked schedule generally made smaller changes (mean = 140 ms following no KR; 221 ms following KR) than participants with a random schedule (mean = 170 ms following no KR; 247 ms following KR).

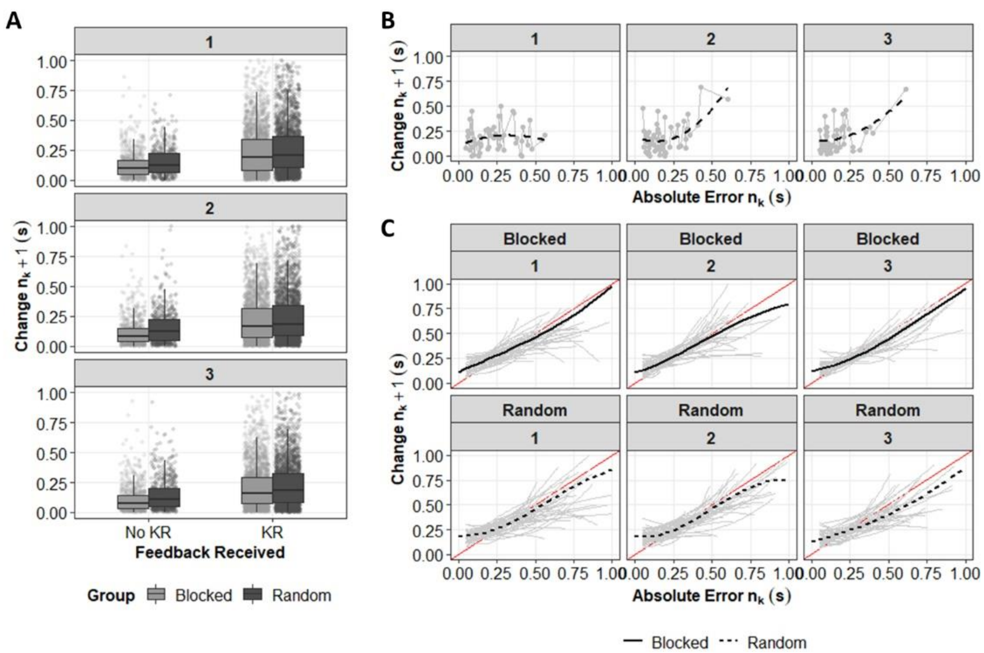

Figure 5. (A) Performance change on the subsequent trial as a function of group, block, and knowledge of results (KR) on the previous trial. (B) Example data and cubic fits are shown for a single participant with a random practice schedule. Performance change on the subsequent trial is shown as a function of absolute error on the previous trial. (C) Predictions from the cubic mixed-effects model are shown as thick black (block scheduled participants) and dashed lines (randomly scheduled participants). Thin grey lines show the best fitting curves for individual participants. Predicted change is shown as a function of group, absolute error on the previous trial, and set of trials (1-3). A thin red diagonal line with an intercept of 0 and slope of 1 shows a proportional corrections of the same magnitude as the previous error.

Changes in performance following KR. Focusing on only those trials following the receipt of KR, we modeled the relationship between the change in performance and absolute error on the previous trial of the same target as shown in Figure 5B. The best fitting model was a cubic polynomial (full details are given in Supplemental Appendix i). Traces for each individual participant are shown as thin grey lines in Figure 5C, with the predictions of the mixed-effect model shown as thick colored lines. Critically, in the mixed-effect model, there were statistically significant linear, $F(1,730.2)=9.96$, $p=0.002$, quadratic, $F(1,2446.7)=17.19$, $p<0.001$, and cubic effects of Lag AE, $F(1,1978.4)=6.77$, $p=0.010$. There was also a statistically significant main-effect for Group, $F(1,152)=9.78$, $p=0.002$, showing that the groups differed in their intercepts. However, these effects were superseded by a Group × Set × Lag AE interaction for the linear effect, $F(2, 11639.9)=3.68$, $p=0.025$. No higher order interactions were significant for the quadratic (p's>0.050) or cubic effects (p's>0.138).

Qualitatively, this interaction is illustrated in Figure 5C; quantitatively, we can understand this interaction by solving for the predicted change in performance at different magnitudes of Lag AE. For instance, given a Lag AE = 500 ms, block scheduled participants were estimated to change their response by 462 ms in Set 1, 470 ms in Set 2, and 443 ms in Set 3. Randomly scheduled participants were estimated to change their response by 457 ms in Set 1, 460 ms in Set 2, and 404 ms in Set 3. Thus, following a 500 ms error, block scheduled participants tended to make more proportional changes on the subsequent trial. Similarly, given a previous absolute error of 0 ms, block scheduled participants were estimated to change their response by 110 ms in Set 1, 106 ms in Set 2, and 122 ms in Set 3. Randomly scheduled participants were estimated to change their response by 185 ms in Set 1, 186 ms in Set 2, and 131 ms in Set 3. Thus, randomly scheduled participants were more likely to erroneously change their performance following a good performance (e.g., Lag AE = 0), and to not proportionally adjust their performance an error (e.g., Lag AE = 500).

Associations with Long Term Learning

From the analyses thus far, data suggest that practice schedules have a significant effect on determinants, how much participants change their responses following no KR, and the relationship between change and previous errors. Note that the intercept from the mixed model is conceptually the same as the change following no KR, but we use the empirically-observed change following no KR because that reflects a “correct” trial in mind of a participant, whereas the intercept of the mixed model reflects a hypothetical perfect trial. Beyond these group level differences, however; it is important to understand how these variables relate to learning on an individual level. These regression results are summarized in Figure 6 and presented fully in Supplemental Appendix i. There were statistically significant differences between groups on the retention test, controlling for any of the other variables (p’s<0.008). There were also statistically significant differences between groups on all three of these variables, controlling for error on the retention test (p’s<0.030). However, there was not a statistically significant relationship between the determinant and the retention test (p=0.672). And although the other two variables showed statistically significant relationships with learning (p’s<0.029), they were in the opposite direction of the group effect. This incongruence between the group-level pattern and the individual-level pattern makes these variables cases of Simpson’s paradox (Kievit, Frankenhuis, Waldorp, & Borsboom, 2013), and suggests that these variables cannot immediately explain the learning effect.

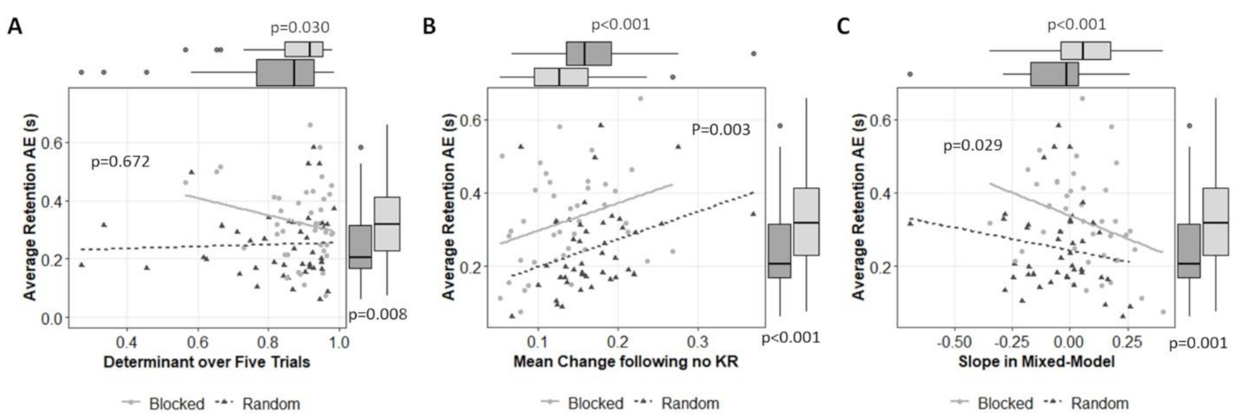

Figure 6. Scatterplots showing the relationship average absolute error (AE) on the retention test as a function of: (A) the determinant of the correlation matrix in target space; (B) the mean change following no-KR trials; and (C) the linear slope from the cubic mixed model. *P*-values in the margins reflect the difference between groups, controlling for the other variable (e.g., p=0.030 reflects the group difference in determinants controlling for retention AE; p=0.008 reflects the group difference in retention AE controlling for the determinant). *P*-values embedded in the scatterplot reflect the relationship between the two variables controlling for group.

DISCUSSION

In this study, we thought that examining how participants adjust from trial to trial might yield insights into the contextual interference effect. A novel contribution of our work is looking at sequential effects in both trial space (e.g., the previous trial in the absolute order they happened) and target space (e.g., the last trial of the same target). We expected that random practice schedules would invoke forgetting and *reconstruction* processes (e.g., Lee & Magill, 1983; 1985), which would be evident in a positive correlation between errors in target space. Additionally, we thought that random practice schedules could lead to *elaboration* processes (e.g., Shea & Zimny, 1983; 1988), which would be evident in more adaptive responses to KR, especially later in practice when the different targets are clearly distinguished from each other.

Our first hypothesis was supported; random practice schedules were associated with positive correlations between responses in target space, but not in trial space. In contrast, the blocked practice group showed very little difference in correlations between trial space and target space, and those correlations were all quite small (to nil). For random practice participants, these correlations were small but reliably positive (r's between 0.10 to 0.15), making them notably larger than the correlations in either phase space and larger than blocked practice participants (r's between 0.00 to 0.05).

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 target space. First, examining responses following "correct" no-KR trials and incorrect trials with KR, we found that participants changed their responses more following KR trials than no-KR trials (replicating Lee & Carnahan, 1990). This is positive adaptive behavior; the most appropriate action following a correct trial is to do the same thing again, whereas the most appropriate action following an error is to change one's behavior (Haith & Krakauer, 2013; Sutton & Barto, 2018). However, we did not find evidence that the degree of this difference depended on participants' practice schedule (i.e., no significant KR × Group interaction). Second, we focused our analysis on only those trials following KR to see how participants responded to error feedback. In that analysis, we did find statistically significant differences in the way the block- and random-practice groups responded to errors, but in a manner opposite to our predictions. Specifically, we found that following more ‘correct’ trials, blocked participants made smaller changes on the subsequent trial, and that following more errorful trials blocked participants made changes that were proportional to the previous error. Although these mean differences were small (perhaps due to the difficult nature of this task), these findings are counter to what we predicted from the elaboration and distinctiveness hypotheses.

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. For random practice, in contrast, the individual is forced to forget the action plan because they must move on to a different trial, requiring reconstruction of the action plan the next time that stimulus is observed (Lee & Magill, 1983; 1985). In the present study, participants who completed random practice schedules appear to be using both the memory of their last response (reflected in positive correlations), plus the feedback they received (reflected in adaptive changes from trial to trial), 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 (reflected in trivial correlations) but can use feedback from trial to trial more effectively (reflected in more adaptive changes). Thus, we see something of a “response inertia” in the random practice participants, who move closer to the target over time but are slow to adapt (i.e., overshoots are 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 yet 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 moderated long-term learning with a blocked schedule, but not a random schedule. Specifically, in the blocked practice group, individuals with worse working memory showed better retention. 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 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 yielding some complementary new insights. Specifically, our data reinforce the argument 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 also extend this past-work, showing the different relationships between consecutive errors in both trial 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 the learner sees a stimulus of the same target (Lee & Magill, 1983; 1985).

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 were no *a priori* power calculations for the myriad statistical tests we conducted in this secondary analysis.

Additionally, although we saw large group differences in the determinant, change following no-KR trials, and the slope of the mixed models in target space, we did not find the same pattern at an individual level (summarized in Figure 6). For the determinant, we simply did find evidence of a relationship between the determinant and long-term learning at the individual level. For change following no-KR trials and slope of the mixed-model, we found that the pattern *reversed* (Kievit et al., 2013). Focusing on change following no KR, at the group-level random practice was associated with *larger* changes following correct performance and better retention test performance (Figure 6B). At the individual-level, however, individuals who had *smaller* changes following correct feedback tended to have better retention test performance. Thus, random practice schedules do lead to better learning, but it does not seem that practice schedules lead to better learning *because* they lead to more adaptive changes.

More work is required to unpack these relationships, but our findings suggest that a simple causal model is not correct (random practice =/= more adaptive change =/= better retention test performance). Within the same practice schedule, however, it is fair to say that individuals who made smaller changes after correct trials showed superior retention.

Similarly, we face a major validity issue if we think about the determinant, intercept, or the slope 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 of internal model. The current results are promising and suggest there is *some* meaningful association between practice schedules, sequential corrections, and learning, but we do not think the models present here are necessarily *the* way to operationalize this research question in future studies.

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 more adaptive 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

Adams, J. A. (1968). Response feedback and learning. *Psychological Bulletin*, *70*, 486.

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.

Bjork, R. A. (1988). Retrieval practice and the maintenance of knowledge. *Practical aspects of memory: Current research and issues*, *1*, 396-401.

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.

Fox, J., & Weisberg, S. (2018). *An R companion to applied regression*. Sage publications.

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.

Kievit, R. A., Frankenhuis, W. E., Waldorp, L. J., & Borsboom, D. (2013). Simpson's paradox in psychological science: a practical guide. *Frontiers in Psychology*, *4*, 513.

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., & Carnahan, H. (1990). Bandwidth knowledge of results and motor learning: More than just a relative frequency effect. *The Quarterly Journal of Experimental Psychology*, *42*(4), 777-789.

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.

Li, Y., & Wright, D. L. (2000). An assessment of the attention demands during random-and blocked-practice schedules. *The Quarterly Journal of Experimental Psychology Section A*, *53*(2), 591-606.

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. (2019). 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

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

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

Roediger III, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. Trends in cognitive sciences, 15(1), 20-27.

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

Schmidt, R. A., & Lee, T. D. (2011). *Motor control and learning: A behavioral emphasis* (5th Ed.). Human kinetics.

Schmidt, R. A., Young, D. E., Swinnen, S., & Shapiro, D. C. (1989). Summary knowledge of results for skill acquisition: support for the guidance hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(2), 352.

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.

Shea, J. B., & Zimny, S. T. (1983). Context effects in memory and learning movement information. *Advances in Psychology*, *12*, 345-366.

Shea, J. B., & Zimny, S. T. (1988). Knowledge incorporation in motor representation. In *Advances in psychology* (Vol. 50, pp. 289-314). North-Holland.
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.
Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press.
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
Thorndike, E. L. (1927). The law of effect. The American Journal of Psychology, 39(1/4), 212-222.
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.
Winstein, C. J., & Schmidt, R. A. (1990). Reduced frequency of knowledge of results enhances motor skill learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(4), 677.
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

Output 1. Model of mean change as a function of group, previous KR-type, and block.

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method [
lmerModLmerTest]
Formula: mean_change ~ group * lag_KR * block + (1 | participant) + (1 |
    block:participant) + (1 | lag_KR:participant)
   Data: ACQ_by_KR
Control: lmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 5e+05))

REML criterion at convergence: -1447.9

Scaled residuals:
     Min       1Q   Median       3Q      Max
-2.43725 -0.50361 -0.03948  0.37904  3.06978

Random effects:
 Groups             Name        Variance  Std.Dev.
 block:participant  (Intercept) 0.0004213 0.02053
 lag_KR:participant (Intercept) 0.0002305 0.01518
 participant        (Intercept) 0.0026947 0.05191
 Residual                       0.0014861 0.03855
Number of obs: 504, groups:
block:participant, 252; lag_KR:participant, 168; participant, 84

Type III Analysis of Variance Table with Satterthwaite's method
                   Sum Sq Mean Sq NumDF DenDF  F value    Pr(>F)
group              0.00741 0.00741     1    82   4.9837 0.0283135 *
lag_KR             0.52979 0.52979     1    82 356.5003 < 2.2e-16 ***
block              0.02994 0.01497     2   164  10.0740 7.473e-05 ***
group:lag_KR       0.00056 0.00056     1    82   0.3738 0.5426163
group:block        0.00052 0.00026     2   164   0.1750 0.8396046
lag_KR:block       0.02741 0.01371     2   164   9.2225 0.0001601 ***
group:lag_KR:block 0.00137 0.00068     2   164   0.4597 0.6322625
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Output 2. Comparison of linear, quadratic, and cubic random-effects in unconditional models to determine the best fitting "shape" of the Lag AE variable.

```
Models:
RE_mod_CHANGE_linear: target_absolute_change ~ 1 + target_lag_absolute_error + (1 +
target_lag_absolute_error | participant) + (1 | block)

RE_mod_CHANGE_quad: target_absolute_change ~ 1 + target_lag_absolute_error +
I(target_lag_absolute_error^2) + (1 + target_lag_absolute_error + I(target_lag_absolute_error^2)
| participant) + (1 | block)

RE_mod_CHANGE_cube: target_absolute_change ~ 1 + target_lag_absolute_error +
I(target_lag_absolute_error^2) + I(target_lag_absolute_error^3) + (1 + target_lag_absolute_error
+ I(target_lag_absolute_error^2) | participant) + (1 | block)
                     npar     AIC     BIC logLik deviance   Chisq Df       Pr(>Chisq)
RE_mod_CHANGE_linear    7 -8175.8 -8123.3 4094.9  -8189.8
RE_mod_CHANGE_quad     11 -8251.9 -8169.3 4137.0  -8273.9 84.0540  4       < 2.2e-16 ***
RE_mod_CHANGE_cube     12 -8258.9 -8168.8 4141.5  -8282.9  9.0118  1       0.002682 **
```

* Note that models failed to converge with a random cubic slope, so that term was dropped from the model. Also, a quartic model (not shown) had a worse AIC than the cubic model (indicating a risk of overfitting). Therefore, the model with cubic fixed effects and quadratic random effects was carried forward for all subsequent models.

Output 3. Model regressing change in performance onto block, group, previous absolute error (in target space) and the interactions of those variables.

```
linear mixed model fit by REML. t-tests use Satterthwaite's method [
lmerModLmerTest]
Formula: target_absolute_change ~ block * group * target_lag_absolute_error +
    block * group * I(target_lag_absolute_error^2) + block *
    group * I(target_lag_absolute_error^3) + (1 + target_lag_absolute_error +
    I(target_lag_absolute_error^2) | participant) + (1 | Target)
   Data: ACQ4
Control: lmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 5e+05))

REML criterion at convergence: -8237.7

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.4406 -0.6490 -0.1607  0.5053  5.8848

Random effects:
 Groups      Name                           Variance  Std.Dev. Corr
 participant (Intercept)                    2.842e-03 0.053307
             target_lag_absolute_error      6.731e-02 0.259445 -0.72
             I(target_lag_absolute_error^2) 1.240e-01 0.352120  0.39 -0.80
 Target      (Intercept)                    3.225e-05 0.005679
 Residual                                   3.099e-02 0.176049
Number of obs: 13542, groups:  participant, 84; Target, 3:

Type III Analysis of Variance Table with Satterthwaite's method
                                            Sum Sq Mean Sq NumDF   DenDF F value    Pr(>F)
block                                      0.06614 0.03307     2 12945.1  1.0671  0.344043
group                                      0.30324 0.30324     1   152.0  9.7842  0.002110 **
target_lag_absolute_error                  0.30859 0.30859     1   730.2  9.9566  0.001669 **
I(target_lag_absolute_error^2)             0.53264 0.53264     1  2446.7 17.1858 3.505e-05 ***
I(target_lag_absolute_error^3)             0.20692 0.20692     1  1978.4  6.6762  0.009842 **
block:group                                0.19920 0.09960     2 12943.0  3.2136  0.040243 *
block:target_lag_absolute_error            0.06715 0.03357     2 11638.2  1.0832  0.338529
group:target_lag_absolute_error            0.24209 0.24209     1   730.1  7.8110  0.005329 **
block:I(target_lag_absolute_error^2)       0.12693 0.06346     2  9370.2  2.0476  0.129096
group:I(target_lag_absolute_error^2)       0.11623 0.11623     1  2446.4  3.7503  0.052914 .
block:I(target_lag_absolute_error^3)       0.15249 0.07624     2  6612.8  2.4600  0.085513 .
group:I(target_lag_absolute_error^3)       0.07041 0.07041     1  1978.0  2.2718  0.131907
block:group:target_lag_absolute_error      0.22804 0.11402     2 11639.9  3.6788  0.025282 *
block:group:I(target_lag_absolute_error^2) 0.18575 0.09287     2  9374.2  2.9966  0.050004 .
block:group:I(target_lag_absolute_error^3) 0.12271 0.06135     2  6616.8  1.9796  0.138207
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Output 4. Predicting the determinant in target space as a function of group and average AE on the retention test.

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

Residuals:
     Min       1Q   Median       3Q      Max 
-0.55131 -0.04763  0.04533  0.08842  0.16994 

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)    
(Intercept)       0.86723    0.03722  23.298   <2e-16 ***
rand.c           -0.06867    0.03100  -2.215   0.0296 *  
ave_ae_Retention -0.05069    0.11923  -0.425   0.6719    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1361 on 81 degrees of freedom
Multiple R-squared:  0.05764, Adjusted R-squared:  0.03437 
F-statistic: 2.477 on 2 and 81 DF,  p-value: 0.09031
```

Output 5. Predicting average AE on the retention test as a function of group 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

Residual standard error: 0.1267 on 81 degrees of freedom
Multiple R-squared:  0.08388, Adjusted R-squared:  0.06126 
F-statistic: 3.708 on 2 and 81 DF,  p-value: 0.02877
```

Output 6. Predicting average change following “correct” feedback (within the 50-ms bandwidth) as a function of group and average AE on the retention test.

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

Residuals:
     Min       1Q   Median       3Q      Max
-0.12616 -0.02769 -0.00352  0.02553  0.18937

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)       0.11293    0.01388   8.137 4.04e-12 ***
rand.c            0.04220    0.01156   3.651 0.000461 ***
ave_ae_Retention  0.13410    0.04445   3.017 0.003414 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.05073 on 81 degrees of freedom
Multiple R-squared:  0.1782,  Adjusted R-squared:  0.1579
F-statistic: 8.784 on 2 and 81 DF,  p-value: 0.0003527
```

Output 7. Predicting average AE on the retention test as a function of group and the average change following “correct” feedback (within the 50-ms bandwidth).

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

Residuals:
     Min       1Q   Median       3Q      Max 
-0.20515 -0.08808 -0.02392  0.06408  0.34403 

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)    
(Intercept)          0.17231    0.03999   4.309 4.58e-05 ***
rand.c              -0.09864    0.02745  -3.594 0.000558 ***
`mean_Change_No KR`  0.75315    0.24967   3.017 0.003414 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1202 on 81 degrees of freedom
Multiple R-squared:  0.1746,  Adjusted R-squared:  0.1542 
F-statistic: 8.566 on 2 and 81 DF,  p-value: 0.000422
```

Output 8. Predicting the individual slope from the mixed-model (i.e., the proportionality of correction) as a function of group and average AE on the retention test.

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

Residuals:
     Min       1Q   Median       3Q      Max 
-0.61986 -0.10627  0.02593  0.10862  0.26761 

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)    
(Intercept)       0.09134    0.04415   2.069 0.041758 *  
rand.c           -0.13373    0.03677  -3.637 0.000483 ***
ave_ae_Retention -0.31446    0.14142  -2.224 0.028957 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1614 on 81 degrees of freedom
Multiple R-squared:  0.1541,	Adjusted R-squared:  0.1332 
F-statistic: 7.377 on 2 and 81 DF,  p-value: 0.00113
```

Output 9. Predicting average AE on the retention test as a function of group and the individual slope from the mixed-model (i.e., the proportionality of correction).

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

Residuals:
     Min       1Q   Median       3Q      Max 
-0.18873 -0.08205 -0.02115  0.07293  0.33567 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.28652    0.01344  21.325  < 2e-16 ***
rand.c      -0.09456    0.02836  -3.334  0.00129 ** 
slope.c     -0.18295    0.08228  -2.224  0.02896 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1231 on 81 degrees of freedom
Multiple R-squared:  0.1347,	Adjusted R-squared:  0.1133 
F-statistic: 6.303 on 2 and 81 DF,  p-value: 0.002857
```

Supplemental Appendix ii. Mixed-effect regressions and correlations with long-term learning, presented in the original pre-print.

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).

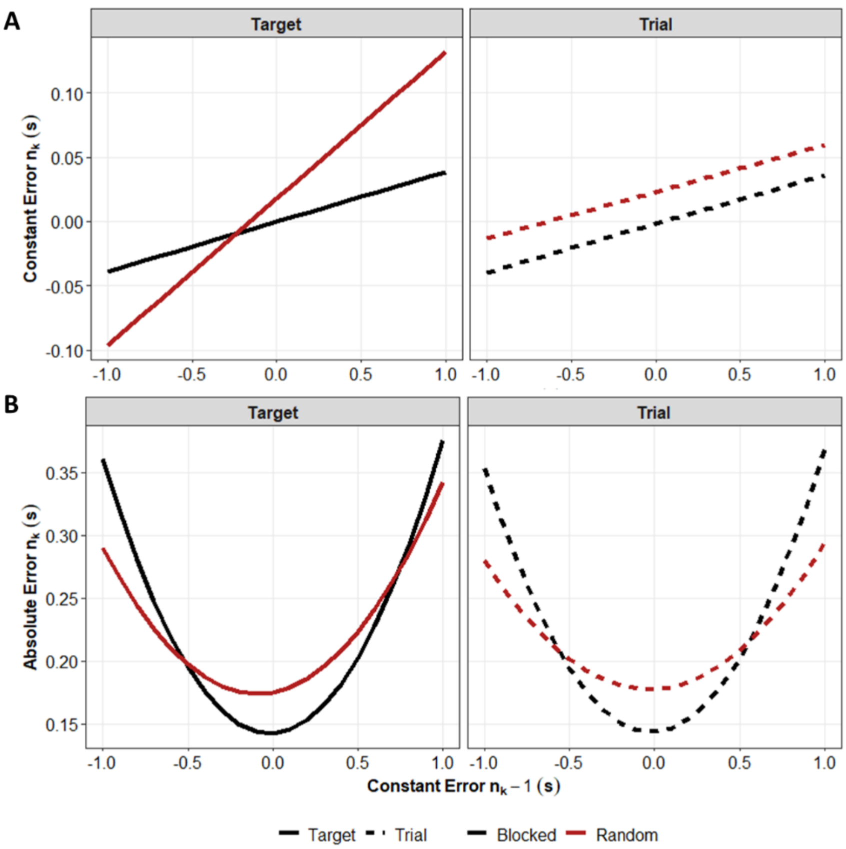

Figure 5. The model predictions for constant error on the next trial (A) or absolute error on the next trial (B) as a function of the previous constant error. Coefficients for all of the models are provided in the supplemental appendix. Solid lines indicate predictions from the model in target space,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

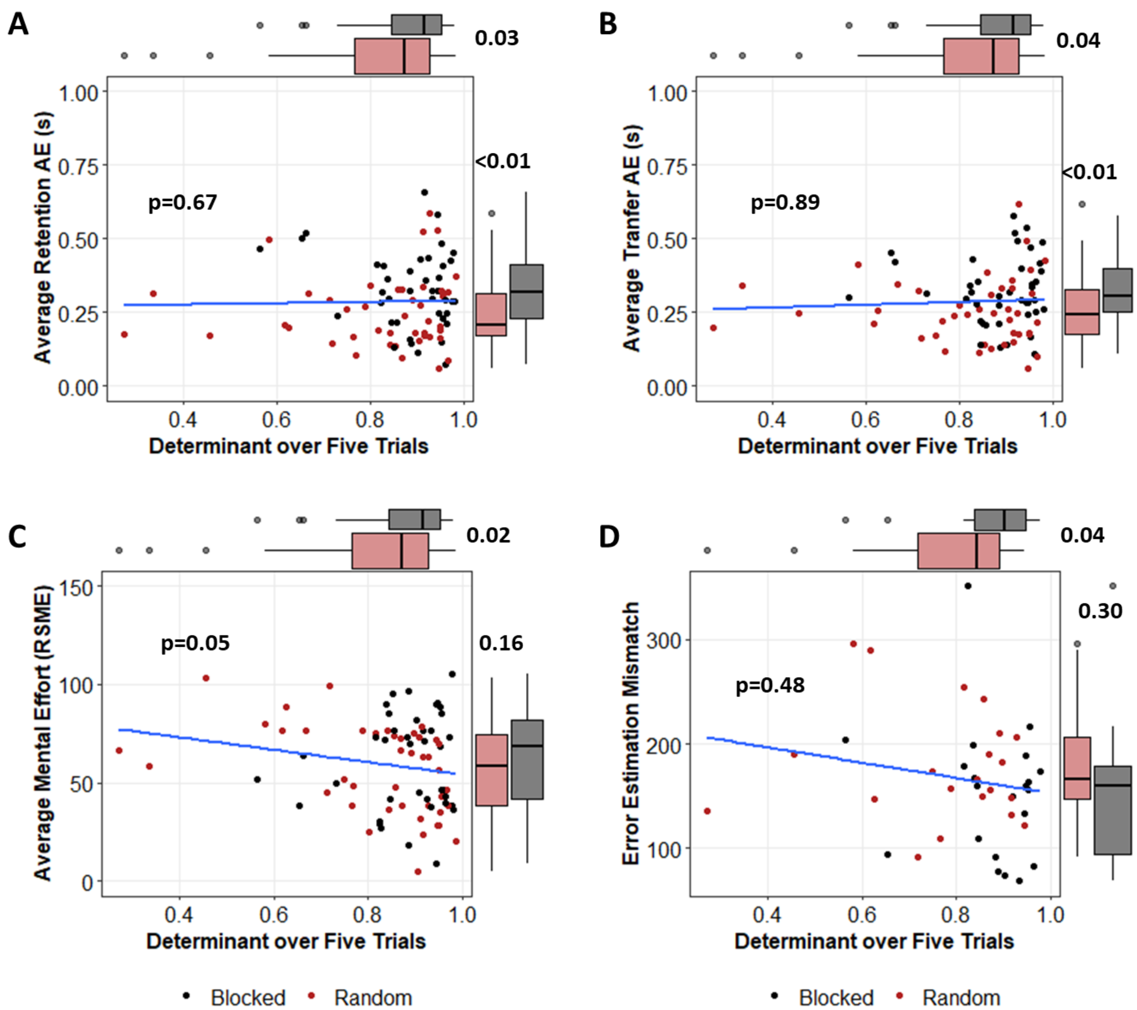

Figure 6. The average absolute error (AE) during retention (A) and transfer tests (B), plus the average from the rating scales of mental effort (RMSE; C), and the mis-match between actual error and estimated error (D) as a function of the determinant in target space and group. *P*-values are given in 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.