

Performance-Based Adaptive Schedules Enhance Motor Learning

Younggeun Choi

Department of Biokinesiology and Physical Therapy
Department of Computer Science
University of Southern California, Los Angeles

Feng Qi

Department of Biokinesiology and Physical Therapy
Department of Neuroscience
University of Southern California, Los Angeles

James Gordon

Nicolas Schweighofer

Department of Biokinesiology and Physical Therapy
University of Southern California, Los Angeles

ABSTRACT. Although investigators have shown that random scheduling of several tasks enhances learning more than blocked scheduling does, the advantages of random scheduling may be limited because it does not take into account the nominal difficulty of each task, the difference in difficulty between tasks, and the skill level of the learner in that type of schedule. The authors propose 2 new algorithms for adaptively determining the nominal difficulty and the number of trials for each task on the basis of both current and delayed performance of the learner ($N = 48$). The authors tested the adaptive algorithms in a 2×2 factorial design, and they show that the algorithms outperform random scheduling when performance is measured on a delayed retention test.

Keywords: adaptive schedules, delayed retention, nominal task difficulty, random scheduling, skill level

Learning of multiple motor skills is ubiquitous in activities such as sports, music, professional skill development, and rehabilitation after brain injury. Although individual skills can be scheduled sequentially, a robust finding in motor learning research is the contextual interference effect: Compared with sequential, or blocked, scheduling, random scheduling of several tasks enhances performance, as measured in delayed retention tests (Boyce & Del Rey, 1990; Carlson, Sullivan, & Schneider, 1989; Carlson & Yaure, 1990; Goode & Magill, 1986; Osu, Hirai, Yoshioka, & Kawato, 2004; Shea & Morgan, 1979; for opposing views, see, e.g., Albaret & Thon, 1998; Brydges, Carnahan, Backstein, & Dubrowski, 2007; Pollatou, Kioumourtzoglou, Agelousis, & Mavromatis, 1997). The theoretical foundations of the effects of random schedules can be traced to Schmidt's (1975) schema theory. Although Schmidt's theoretical predictions of a positive effect of constant versus variable practice schedules on learning have been equivocal (Lee, Magill, & Weeks, 1985; Shapiro & Schmidt, 1982),

when researchers randomize variable practice rather than block it, that practice schedule is effective in improving learning (Lee et al.).

Despite its successes, however, random task scheduling suffers from at least two limitations. First, it does not account for the two components of task difficulty (Guadagnoli & Lee, 2004): (a) the difficulty of the task itself (nominal task difficulty), which is learner-independent, and (b) the skill level of the learner (functional task difficulty), which improves with learning. Second, it does not account for differences in difficulty between tasks: All tasks being learned are treated equally for all learners at all times.

Matching task difficulty to the learner's skill level, both initially and as learning progresses, has the potential to enhance the learning effectiveness of each task. On the one hand, the tasks should not be too easy to perform because mere repetitions do not lead to change in performance. Nor do they seem to induce cortical reorganization, which is produced by the mastering of challenging tasks (Kleim et al., 2002; Nudo, Wise, SiFuentes, & Milliken, 1996; Plautz, Milliken, & Nudo, 2000). On the other hand, when initial difficulty is too large, that is, when the combined high nominal and functional task difficulties create large errors, learning can fail to occur because of the unavailability of appropriate training examples (Sanger, 2004). Task difficulty should be adapted to the learner. Vygotsky (1978) proposed that individuals should maintain task difficulty near the optimal challenge point (or the just-right challenge; Ayres, 1972) to enhance learning effectiveness. In artificial neural network

Correspondence address: Nicolas Schweighofer, Department of Biokinesiology and Physical Therapy, University of Southern California, 1450 E. Alcazar Street, Los Angeles, CA, 90089, USA. E-mail address: schweigh@usc.edu

studies, adaptive scheduling of task difficulty has been shown to enhance learning of motor tasks. Sanger (1994) showed that training a neural network to learn the dynamics of a multijoint arm was much faster when the speed of desired movements was slowly increased than when the network was trained on fast movements. Ivanchenko and Jacobs (2003) tested Bernstein's (1967) hypothesis that developmental progression accelerates motor learning and showed that neural network training regimens that included developmental progression outperformed fixed training on motor tasks that were relatively difficult to learn.

Taking into account the difference in difficulty between tasks has the potential to enhance learning performance globally. The total amount of practice is the single most important variable for skill acquisition (Schmidt & Lee, 1999). Furthermore, negatively accelerated monotonically increasing functions, such as exponentials or power functions, model performance improvement as a function of practice trials well (Liu, Mayer-Kress, & Newell, 2003; Schmidt & Lee). According to such models, in each new practice trial, the gain in performance is less than the gain in the previous trial; at the limit, the gain tends toward zero. Moreover, depending on nominal difficulty and the skill of the learner, the rate of performance improvement varies from task to task. Thus, in simple random scheduling, each new trial of an easy task may induce a labor in vain effect (Nelson & Leonesio, 1988) as performance improvement levels off. For a harder task, however, performance still may be relatively poor but still improving. Because artificial neural networks face similar issues, researchers have devised active input selection methods to select inputs (tasks) that potentially lead to large decrements in error (Zhang, 1994) or are maximally informative (MacKay, 1992). Those methods, which de facto implement adaptive scheduling in the number of trials, largely reduce the training time of the neural networks (see, e.g., Zhang).

Thus, the results of previous research have suggested that methods that dynamically maintain performance near an optimal challenge point at which performance gains during learning are maximal may improve human learning of a motor task. Furthermore, methods that dynamically allocate learning trials differentially among tasks, with fewer trials allocated to easy tasks and more trials allocated to hard tasks, may improve global learning performance in multi-task situations. In the present study, we specifically tested whether a multitask motor-training program that included adaptive practice schedules both in difficulty and in the number of trials for each task could be more effective than random scheduling in enhancing delayed performance.

Method

Participants and Design

Participants were 48 college students (18 men and 30 women, aged 26.0 ± 1.3 years [$M \pm SD$]). We conducted the experiment over 4 consecutive days. Participants provided informed consent, and the Human Subjects Institutional

Review Board of the University of Southern California approved the study protocol.

Participants had to learn four visuomotor transformations over the course of three learning sessions, each separated by 24 hr. A learning session consisted of 120 trials. On each trial, we instructed participants to move a cursor shown on a computer screen from an initial position to a target within a limited time, the allocated movement time (*AMT*), by using a type of force-feedback joystick—a spaceball (HP 5000). A color-coded target and an associated angular relationship between spaceball movement and cursor movement defined each visuomotor transformation. We randomly assigned the four angles (-30° , 60° , -90° , and 120°) to the four color-coded targets for each participant. The *AMT* to reach the target controlled nominal difficulty for each task.

We randomly assigned participants to one of four conditions (12 participants per condition) in a 2×2 factorial design: (a) fixed number of trials and fixed difficulty (Fix), (b) adaptive number of trials and fixed difficulty (AdapTr), (c) adaptive difficulty and fixed number of trials (AdapDif), and (d) adaptive number of trials and adaptive difficulty (AdapTrDif). In the Fix and AdapDif conditions, the number of trials was constant: 30 trials per task per session. In the AdapTr and AdapTrDif conditions, we allowed the number of trials per task to vary between 15 and 75 in Sessions 2 and 3. In the Fix and AdapTr conditions, the difficulty was constant: *AMT* = 0.6 s. In the AdapDif and AdapTrDif conditions, we allowed *AMT* to vary between 2.5 and 0.6 s in all sessions, and we set *AMT* at 2.5 s at the beginning of Session 1 for all tasks.

Experimental Procedure

Participants sat comfortably at a table in front of a computer display and held the spaceball in their right hand. We normalized the participants' position both across participants and across days: We measured the distance between the spine (at the C3–C4 level) and the acromion process of the scapula. We positioned participants so that their midline was aligned with the middle of the computer display. We positioned the spaceball and attached it on the table with Velcro at 15 cm in front of the participants' body and to the right of the midline at the spine–scapula distance.

At the beginning of each trial, a colored target was shown as a 3-cm-diameter disk located at 10 cm above the initial position. A white fixation cross indicated the initial position. After 1 s, a change in the color of the fixation cross from white to the color of the target indicated the go signal. After a variable reaction time (RT), participants started to move the spaceball. Participants had *AMT* (seconds) to move the cursor onto the target. The cursor movement was visible on the screen at any time $t \leq AMT$. At $t = AMT$, the cursor movement stopped. The cursor and the target were visible during the intertrial interval, which lasted a minimal 1 s. We adjusted the variable intertrial interval so that the trial duration was always 5.5 s (go signal duration + RT + *AMT* + intertrial interval = 5.5 s). If the RT was larger than

1 s, we excluded the trial, and the screen displayed, “Next time move faster,” until the next trial started.

We defined *performance error* as the distance between cursor position and target position at *AMT*. If the cursor was on target for 100 ms before the end of the trial, then performance error was zero, and we terminated the trial and played a bird-chirping sound. The maximum performance error was 16 cm. Because we manipulated task difficulty to keep performance relatively constant in the adaptive difficulty conditions (see *AdapDif* and *AdapTrDif*), participants would have found it difficult to perceive any progress in performance directly from the task. Thus, to indicate progress in performance, we provided a display above the target that showed the score of each task at all times and in all conditions. The score was given in the same color as the task. We computed performance feedback by scaling *AMT* (see Equation 4) between 0 and 100 (we computed *AMT* in all schedules but used it for feedback only in the *Fix* and *AdapTr* schedules).

Each session began with a pretest and ended with a post-test. There were four pseudorandomly distributed trials per task in each test. A fourth test was given 24 hr after the last session. Note that pretests on Days 2 and 3 and the test on Day 4 can be considered 24-hr delayed retention performance tests.

Fix condition. The *Fix* schedule was the control condition. We scheduled tasks pseudorandomly so that each task occurred once in a block of four trials. Thus, there were 30 trials per task per session, and we kept the difficulty constant at $AMT = 0.6$ s.

AdapTr condition. In this condition, we adjusted the number of trials per task in Sessions 2 and 3 by distributing the total number of trials among the tasks:

$$NoOfTrial(Task) = totalNoOfTrials \times PerfE_{norm}(Task), \quad (1)$$

where *totalNoOfTrials* is the total number of trials in one learning session and $PerfE_{norm}(Task)$ is the normalized performance error for the task.

To obtain an estimate of performance for each task, we measured performance error directly with a test before each training session. However, because we gave this test at fixed (high) difficulty level, it had to be short (or the purpose of adaptive difficulty schedules would be defeated). Thus, because we estimated performance in a few trials, it was somewhat variable. Therefore, we also used past performance error to obtain a more accurate estimate of performance. Although performance during practice is often not a good indicator of long-term retention (Cahill, McGaugh, & Weinberger, 2001; Winstein, 1991), it reflects a measure of task difficulty. We computed normalized performance error by combining those two (imperfect) measures in the following parameter-free equation:

$$PerfE_{norm}(task) = \frac{PerfEc(task) \times PerfEp(task)}{\sum_t [PerfEc(t) \times PerfEp(task)]}, \quad (2)$$

where $PerfEc(task)$ is the performance error obtained at the pretest preceding practice in the current session, which can be considered a 24-hr delayed retention performance test; $PerfEp(task)$ is the performance error in the test immediately following the previous practice session, and the denominator is a normalizing factor with the sum over all tasks.¹ Because tasks compete for the number of trials, we provided a range of number of trials to avoid the possibility that one task would take all the available trials in a session (maximum = 75 trials, minimum = 15 trials). Once we determined the number of trials (after the pretest in Sessions 2 and 3), we scheduled the trials pseudorandomly to minimize consecutive trials of the same task: Consecutive trials could occur only if one task was allocated more than 60 trials.

AdapDif condition. In the adaptive schedules, we attempted to maintain an optimal challenge for each task by manipulating the allocated movement time $AMT(t)$ at each trial, as follows: If the current performance error was above a reference error, then task difficulty was too high, and we increased *AMT*. If, on the contrary, current performance error was below the reference error, then difficulty was too low, and we decreased *AMT*. To accomplish this, we used an error-reduction learning rule, in which we based the change in difficulty on the squared difference between performance error and the reference performance error: $E(t) = (PerfE[t] - PerfE_{ref})^2$, where t is the trial index, $PerfE(t)$ is the performance error on trial t , which is small when performance is good, and $PerfE_{ref}$ is the reference performance error.

We derived our difficulty update algorithm from the assumption that the performance curve of motor learning is relatively well modeled with exponential functions. This model allowed us to compute the derivative of the squared difference between participants' performance and the reference performance error. By using this mathematical minimization technique, we derived a flat performance curve near the reference performance error. At each trial, we adjusted task difficulty $AMT(t)$ by subtracting a fraction of the derivative of the difference between the participant's performance and the reference performance error with respect to the number of trials. Thus, the allocated movement time $AMT(t)$ at trial t is given by

$$AMT(t) = AMT(t-1) - \mu \frac{dE(t)}{dt}, \quad (3)$$

where μ is a small positive constant. By assuming that performance is well modeled by an exponential function of the number of trials t , with the task difficulty controlling the learning rate, $PerfE(t) = A \exp(-AMT[t] \cdot t)$, we obtained the following update rule:

$$AMT(t) = AMT(t-1) \times \{1 + \alpha [PerfE(t) - PerfE_{ref}]\}, \quad (4)$$

where $\alpha > 0$ is a learning rate. Because the optimal challenge point is an unknown parameter, we set the reference performance error $PerfE_{ref}$ at 4 cm after pilot testing. Adaptive

difficulty for each task started at the onset of Session 1 and continued in Sessions 2 and 3. We allowed *AMT* to vary between 2.5 and 0.6 s, and we initially set it at 2.5 s at the beginning of Session 1 for all tasks. We scheduled tasks pseudorandomly as in the Fix condition, with 30 trials per task.

AdapTrDif condition. In this condition, the difficulty is adapted as in the AdapDif condition, and the number of trials is the same as in the AdapTr condition.

Statistical Analysis

For statistical tests on performance error, we first computed the median of the test for each participant and then computed the mean of the median for the group of participants. For all comparisons of means, we first checked the data for normality with the Shapiro–Wilk test and for equal variance with the Levene test. When the data was normally distributed and the variances equal, we used one-way analysis of variance (ANOVA) to analyze the effect of condition on posttest performance. Unless otherwise noted, we used Bonferroni corrections in multiple comparisons. When variances were not equal, we used the robust Brown–Forsythe test for equality of means because this test does not require the assumption of equal variance. In repeated ANOVAs, we used Mauchly’s test to determine sphericity. We computed effect sizes as differences in group means divided by the pooled common standard deviations. Effect sizes between 0.5 and 0.8 are considered moderate, and effect sizes above 0.9 are considered large (Cohen, 1977). We report results as means ± standard errors. Our significance level was $p < .05$.

Results

We report results for the performance error for the average of the four tasks first and then for individual tasks. There was no significant difference between conditions in pretest of Day 1, one-way ANOVA, $F(3, 44) = 5.21, p = .22$. In the delayed retention test on Day 4, there was an effect of condition on performance error; Brown–Forsythe test, $F(3, 22.9) = 4.11, p = .018$. Participants performed better in the AdapTrDif condition than in the control, Fix, condition; performance errors were 2.31 ± 0.42 cm and 5.58 ± 1.18 cm, respectively, $p = .016$ (see Figure 1A). Effect size $ES = 1.11$, which is considered a large effect size. Similarly, participants performed better in the AdapTr condition (2.63 ± 0.42 -cm performance error) than in the Fix condition, $p = .038$ (see Figure 1A). $ES = 0.96$, which is also a large effect size. Although results did not reach significance, we found a trend when we compared performance error in the adaptive difficulty condition (3.45 ± 0.60 cm) with performance error in the control condition, $p = .045$, uncorrected for multiple comparisons, $ES = 0.64$ (see Figure 1A). There was no significant difference for the other comparisons ($p = 1$). Note that in the fixed schedule, some learners could not learn the harder tasks, and they exhibited failure of learning (see Figure 1B, top panel).

One can consider performance error in the Fix schedule on Day 2 as an index of task difficulty (there may have

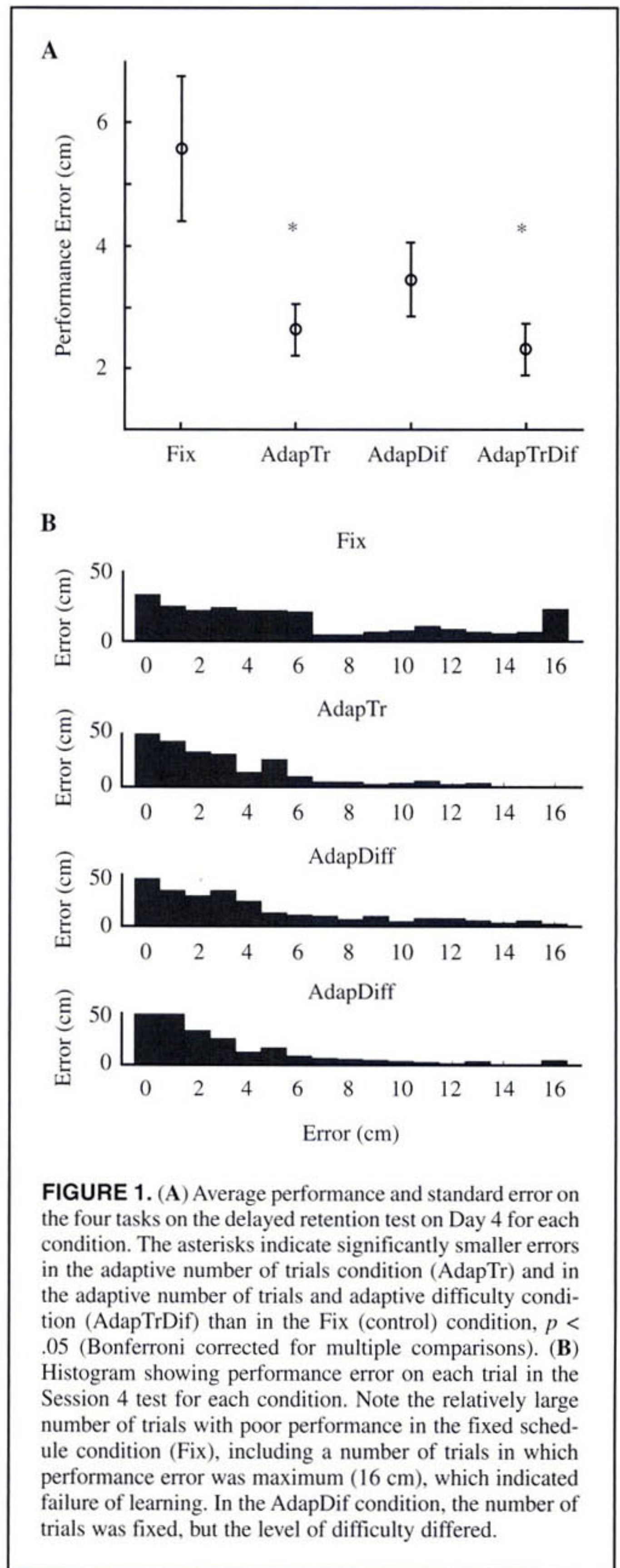


FIGURE 1. (A) Average performance and standard error on the four tasks on the delayed retention test on Day 4 for each condition. The asterisks indicate significantly smaller errors in the adaptive number of trials condition (AdapTr) and in the adaptive number of trials and adaptive difficulty condition (AdapTrDif) than in the Fix (control) condition, $p < .05$ (Bonferroni corrected for multiple comparisons). (B) Histogram showing performance error on each trial in the Session 4 test for each condition. Note the relatively large number of trials with poor performance in the fixed schedule condition (Fix), including a number of trials in which performance error was maximum (16 cm), which indicated failure of learning. In the AdapDif condition, the number of trials was fixed, but the level of difficulty differed.

been floor effects in Days 3 and 4, and performance was very variable in Day 1 because participants were using the joystick for the first time). The task difficulty increased with the rotation angles: Mean performance errors in Day 2 for

Tasks 1 (−30°), 2 (60°), 3 (−90°), and 4 (120°) were 7.90 cm, 8.18 cm, 10.30 cm, and 10.35 cm, respectively.

Figure 2 shows the mean performance across participants for each task and each condition on Day 4. As is apparent in the figure, participants' performance in Tasks 2, 3, and 4 in the Fix schedule tended to worsen in comparison with that in Task 1. As is also apparent in each of the figure panels, the differences in performance between tasks were reduced in the adaptive schedule conditions. Considering all the tasks and schedules on Day 4, there was no main effect of task on performance—repeated ANOVA, $F(3, 132) = 1.85$, $p = .141$ —but there was a significant Task \times Condition interaction effect, $F(9, 132) = 0.26$, $p = .026$. Furthermore, there was no effect of condition for Task 1 (30°), Brown–Forsythe test $p = .10$, or Task 4 (120°), Brown–Forsythe test $p = .259$. However, there were significant effects of condition for Task 2 (60°), Brown–Forsythe test $p = .037$, and Task 3 (−90°), Brown–Forsythe test, $p = .002$.

The AdapTr schedules produced different numbers of trials for each task, and allocated more trials to harder tasks

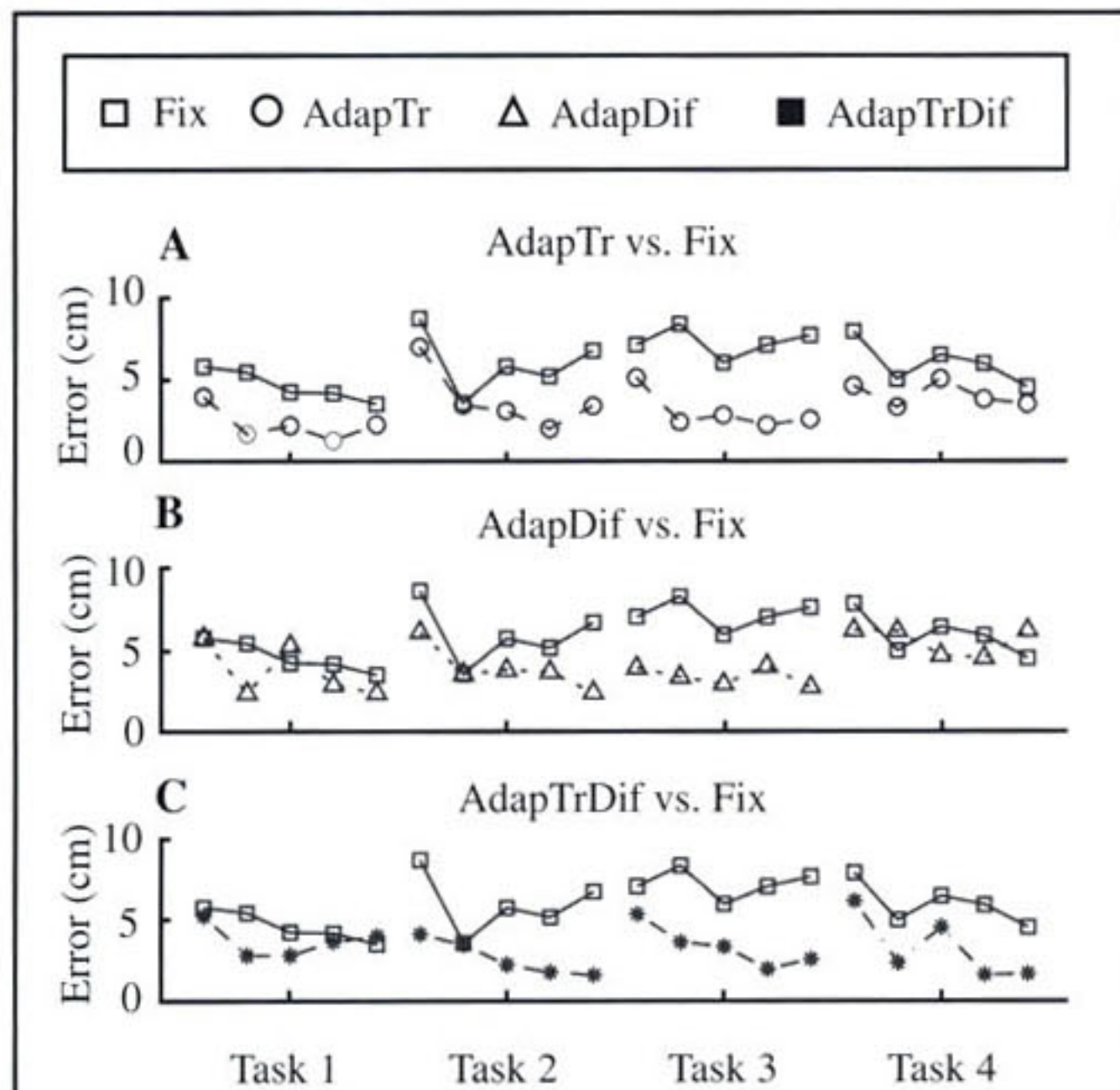


FIGURE 2. Mean performance in the five trials of the delayed retention test given on Day 4 for each task and each schedule. The performance of the fixed schedule (Fix) group is repeated on each panel for comparison with those of the other three adaptive groups. The adaptive groups showed better mean performance (smaller errors) for almost all trials and tasks. The increase in performance was more marked in the two tasks of intermediate difficulty (Tasks 2 and 3) than on the easier task (Task 1) or the more difficult task (Task 4). AdapDif = adaptive difficulty and fixed number of trials, AdapTrDif = adaptive number of trials and adaptive difficulty. In the Fix and AdapDif conditions, the number of trials was constant: 30 trials/task/session. In the AdapTr and AdapTrDif conditions, the number of trials per task varied between 15 and 75 in Sessions 2 and 3. In the Fix and AdapTr conditions, the difficulty was constant; in the AdapDif and AdapTrDif conditions, difficulty varied.

than to easier tasks, as prescribed. Figure 3 shows an example of a task schedule in the AdapTr condition. On Day 2, the mean percentage change in number of trials was 47.7% between Tasks 1 and 2, 87.5% between Tasks 1 and 3, and 67.3% between Tasks 1 and 4. In Day 3, the mean difference in number of trials was 63.5% between Tasks 1 and 2, 67.8% between Tasks 1 and 3, and 69.9% between Tasks 1 and 4. In the AdapTr condition, the number of trials for Task 1 was significantly fewer than 30, which was the number of trials for each task in the Fix condition (t test Day 2, $p = .003$; t test Day 3, $p = .0045$). We then compared the number of trials between tasks in the AdapTr condition. A one-way ANOVA showed a main effect of task, $p = .0064$. There were significantly fewer trials in Task 1 than in all the other tasks on Day 2 and significantly fewer trials in Task 1 than in Task 3 on Day 3, both $ps < .05$.

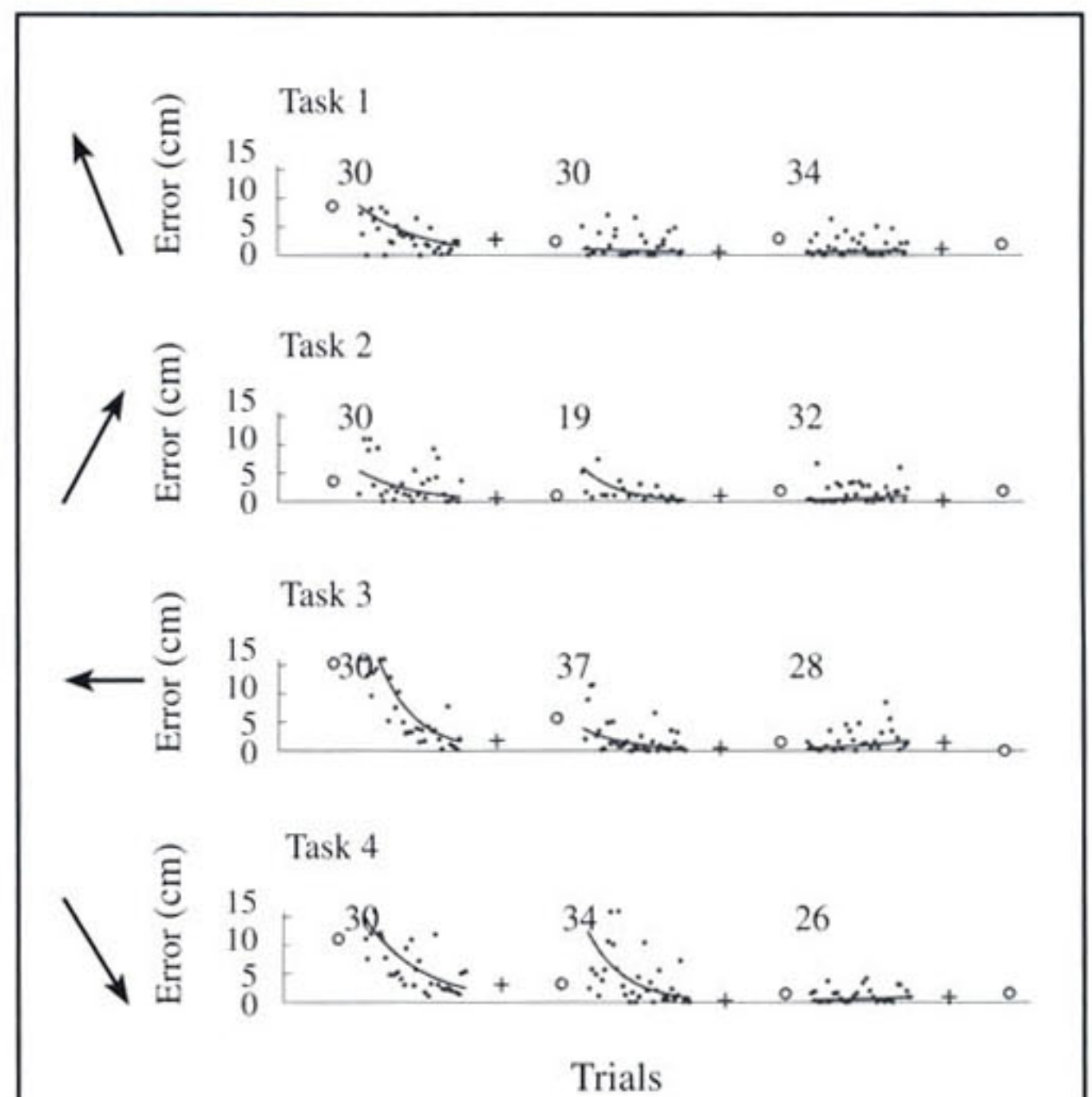


FIGURE 3. Examples of the adaptive number of trials schedule in the AdapTr condition (variable number of trials, constant difficulty) for the four tasks. The arrows on the left indicate the angles of the visuomotor transformation for each task (−30, 60, −90, and 120°). The dots show the performance for each trial for the three practice sessions (from left to right) for Tasks 1–4. Performance error during training has been fitted with exponentials for comparison of the four tasks. Filled circles show median performance in the tests before training. Plus signs indicate median performance in the tests immediately following training. The numbers of trials for each task are shown to the left of performance for each session. For Task 3, performance was relatively poor in the posttest in the pretest in Session 2; thus, a larger number of trials (37) were scheduled for this task in Session 2. On the contrary, this learner's performance was good in Task 2 before Session 2; hence a relatively small number of trials of this task were performed in this session. The total number of trials/session was kept constant at 120 trials.

Although we designed our AdapTr algorithm to minimize the number of consecutive trials, consecutive trials are inevitable if more than half the total number of trials per session (i.e., more than 60 out of 120 trials) are allocated to a single task. The different number of trials for each task between the Fix group and the AdapTr and AdapTrDif groups was potentially confounded with the number of consecutive trials (i.e., if there are more trials for a task, consecutive trials for that task are more likely). Consecutive trials were unlikely to occur in the present experiment, however, because only 2 of the 48 schedules that generated adaptive numbers of trials—12 (participants) \times 2 (adaptive-number-of-trials conditions: AdapTr, AdapTrDif) \times 2 (Day 2, Day 3)—had consecutive trials: one schedule on Day 2 in AdapTr (with 15, 15, 75, and 15 trials per task, respectively) and the other on Day 3 in AdapTrDif (with 17, 20, 15, and 68 trials per task, respectively).

In the adaptive difficulty schedules, participants maintained performance error near the reference performance error, as desired. Figure 4 shows an example of a performance and adaptive difficulty in the AdapDif condition. In the Fix condition, average performances across participants and all 30 trials in Sessions 1 and 2 were 10.44 ± 0.13 cm and 6.75 ± 0.14 cm, respectively. In the AdapDif condition, average performances in Sessions 1 and 2 were 6.73 ± 0.16 cm and 3.84 ± 0.11 cm, respectively. Performance during practice was better in the AdapDif condition than in the

Fix condition in both sessions, two-tailed unbalanced Satterthwaite test, both $ps < .001$. In addition, performance was closer to the reference performance error in the AdapDif condition, $p = .15$, than in the Fix condition, $p < .001$ (one-sample t test with 4 cm as the test value).

In our study, we manipulated task difficulty by allocating movement durations. Thus, adapting the task difficulty was potentially confounded with changing the overall time of practice: The improvement in performance could have simply been caused by an increase in total practice time. To study that possibility, we computed the actual total practice time for the participants in the AdapDif group and correlated it with performance on Day 4. We found no correlation between total practice time and performance in this group, both when we considered all tasks, $r = -.053$, $p = .80$, and when we considered individual tasks: Task 1, $r = .13$, $p = .67$; Task 2, $r = .46$, $p = .13$; Task 3, $r = -.36$, $p = .24$; and Task 4: $r = -.20$, $p = .53$.

Discussion

The results of the present experiment demonstrate that adaptive scheduling can improve individuals' performance in a multisession, multitask learning program. Adapting the number of trials as a function of performance largely improved retention in comparison with maintaining fixed random scheduling. Varying difficulty also improved learning, although to a lesser extent. Furthermore, although

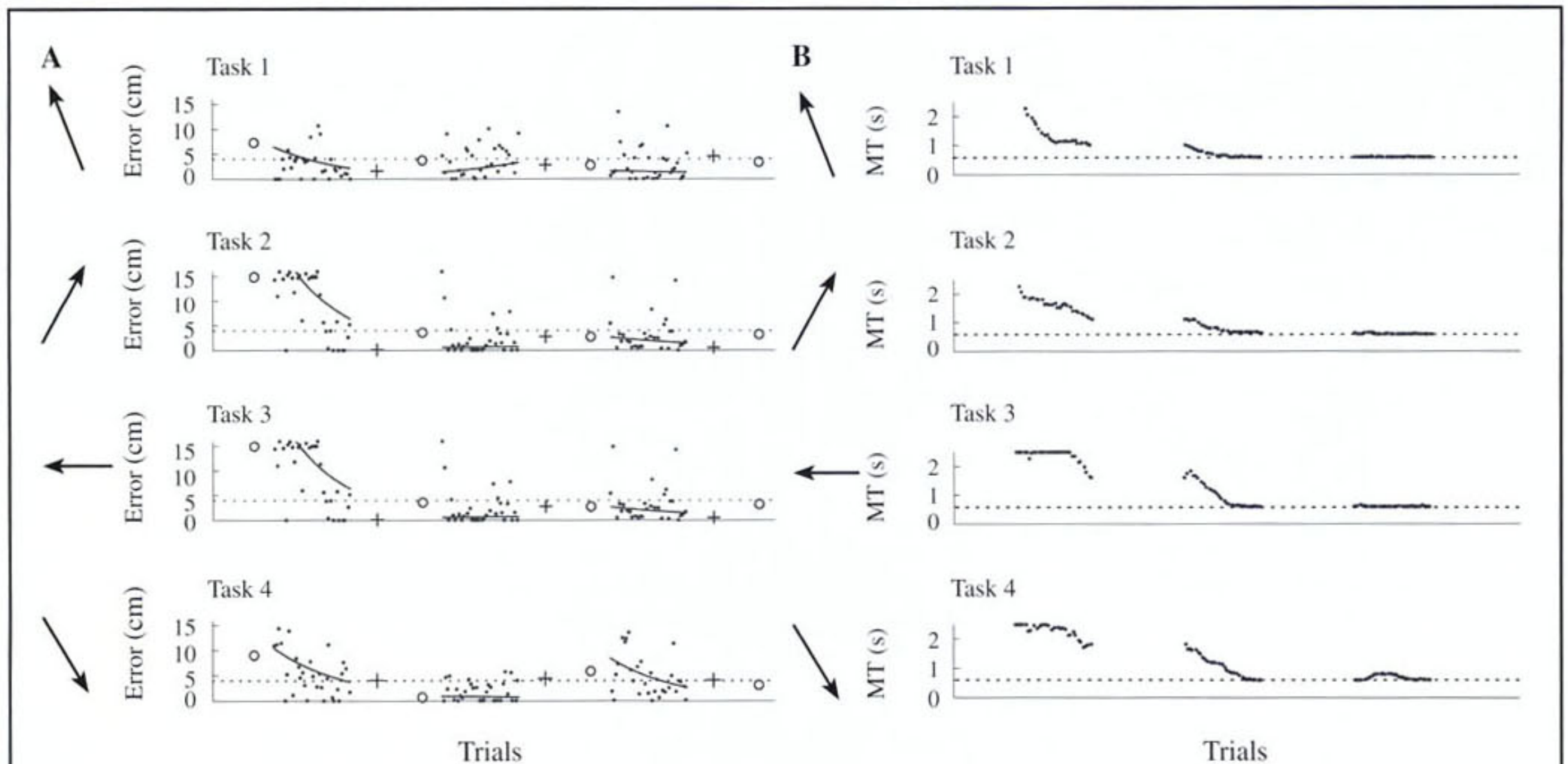


FIGURE 4. Example of the adaptive difficulty schedule for the four tasks. The arrows on the left indicate the angles of the visuo-motor transformation for each task (-30 , 60 , -90 , and 120°). Performance error as a function of trials (cm) is shown in (A). For Tasks 1–4, the average performance error was quickly brought near the challenge point (4 cm). Change in the difficulty parameter, allocated movement time (AMT, in seconds), in each trial is plotted in (B). As performance improved, task difficulty, controlled by AMT, decreased. Note that because Tasks 1 and 2 (upper and upper middle panels) were relatively easy for this participant in Session 1, AMT decreased rapidly in the first session so that he maintained performance around the optimal challenge, as can be seen in the corresponding graphs in A.

performance in the AdapTrDif condition was not significantly different from performance in the AdapTr condition, probably because of a floor effect, it was superior to that in the Fix condition. In all adaptive conditions, performance on all tasks improved. In the Fix schedule, however, some learners could not learn the harder tasks and exhibited failure of learning (Figure 1B). Nominal task difficulty was important in determining positive effects of the adaptive schedules, however: The effects of adaptive schedules were large in tasks of intermediate difficulty, but the schedules had less or no effect on the easiest task, possibly because of a floor effect, and on the most difficult task, possibly because of a slow rate of learning in the adaptive difficulty conditions or a limitation in the allocated number of trials in the adaptive number of trials conditions.

Because of the practical importance of enhancing effectiveness and efficiency in motor learning in a number of domains, such as sports, professional skill development, and rehabilitation, it is surprising that only a few behavioral researchers have examined whether adaptive practice schedules enhance learning. Researchers have designed two types of adaptive scheduling methods, which can be divided into learner-controlled and computer-controlled (as in the present study) methods. In learner-controlled methods, the learner determines the practice schedules. Titzer, Shea, and Romack (1993) showed that a learner-determined schedule had the same beneficial effect as a blocked schedule during acquisition and was equivalent to random practice in retention, thus facilitating both performance and learning. Wu, Magill, and Foto (2005) showed that the performance of a learner-determined-schedule group was superior to that of a yoked control group. A possible limitation of those methods, however, is that the chosen schedules were based on imperfect metacognitive judgments: Because learners suffer from illusions of competence during practice (Simon & Bjork, 2001), those schedules may not optimally enhance long-term retention. Other than our study, we are aware of only one motor learning scheduling study that did not rely on metacognitive judgments but in which the schedule was computer controlled and based on measured performance: the *win-shift/lose-stay method* (Simon, Cullen, & Lee, 2002). In this method, switching to another task occurs only after the learner has achieved a criterion level of success. Although beneficial to learning, a potential limitation of this algorithm is that it cannot distinguish between learning and performance (Cahill et al., 2001; Winstein, 1991) because it uses performance during practice to tailor the schedule. In the method that we propose here, the researcher adapts the number of trials on the basis of immediate and delayed retention performance.

The concept of optimal challenge has been studied in motivation research and has notably been linked to intrinsic motivation. Intrinsic motivation reflects the individual's propensity to engage in a task for its own sake and, in so doing, to seek out and master optimal challenges (i.e., challenges in accordance with the individual's capability; Deci & Ryan, 1985; Hebb, 1955; White, 1959). Intrinsic motivation has been shown to be sustained if the optimal

challenge is itself sustained and if reception of informed feedback shows the individual his or her progress toward the goal (Csikszentmihalyi, 1979). Thus, in future work, researchers should dissect the effects of optimal challenge on enhancing learning (as we have shown here) and on enhancing intrinsic motivation. In further work, researchers should also aim at determining the optimal challenge point meta-parameter. In the present study, we empirically determined and took a reference performance error as an approximate challenge point. Investigators should develop new methods to automatically adjust this parameter, perhaps on the basis of both user preference and performance.

Because different practice schedules yielded large differences in delayed performance error, at least for tasks of intermediate difficulty, we believe that our findings are robust and that they warrant the development of similar methods outside the laboratory. Thanks to the recent availability of relatively cheap and simple motion-capture systems, adaptive task scheduling as described in the present article could be used in real applications, such as sports, or in rehabilitation of hand function after brain injury.

NOTE

1. Our choice of formula for computing normalized performance error was motivated by the *predict* and *correct* steps in the Kalman filter (Welch & Bishop, 2004), in which the Bayesian rule is used: One obtains the posterior estimation of the performance error value by multiplying the previous estimation, $PerfEp(task)$, by the likelihood, $PerfEc(task)$, obtained by current measurements (Thrun, 2000) divided by a normalization term.

ACKNOWLEDGMENTS

National Institutes of Health grant NIH R03 HD050591-02 and National Science Foundation grant IIS 0535282 to Nicolas Schweighofer partially supported this study. The authors thank Usha Guduri for help with programming, Natalie Brainerd and Phu Phuc for help with conducting the experiment, and Carolee Winstein for helpful suggestions.

Biographical Notes

Younggeun Choi is pursuing a doctoral degree in computer science at the University of Southern California. His interests include motor learning schedule and rehabilitation robotics.

Feng Qi is pursuing a doctoral degree in neuroscience at the University of Southern California. His research interest is the cortical neuroplasticity underlying motor learning.

James Gordon is an associate professor and Chair of the University of Southern California Biokinesiology and Physical Therapy Division. He conducts research on motor learning and motor control in healthy participants and patients.

Nicolas Schweighofer directs the Computational Neurorehabilitation and Learning Laboratory at the University of Southern California; his research interests include the behavioral and neural bases of learning and their implications for neurorehabilitation.

REFERENCES

- Albaret, J. M., & Thon, B. (1998). Differential effects of task complexity on contextual interference in a drawing task. *Acta Psychologica (Amsterdam)*, *100*(1-2), 9-24.

- Ayres, A. J. (1972). *Sensory integration and learning disorders*. Los Angeles, CA: Western Psychological Services.
- Bernstein, N. (1967). *The coordination and regulation of movements*. New York: Pergamon Press.
- Boyce, B. A., & Del Rey, P. (1990). Designing applied research in a naturalistic setting using a contextual interference paradigm. *Journal of Human Movement Studies, 18*, 189–200.
- Brydges, R., Carnahan, H., Backstein, D., & Dubrowski, A. (2007). Application of motor learning principles to complex surgical tasks: Searching for the optimal practice schedule. *Journal of Motor Behavior, 39*, 40–48.
- Cahill, L., McGaugh, J. L., & Weinberger, N. M. (2001). The neurobiology of learning and memory: Some reminders to remember. *Trends in Neuroscience, 24*, 578–581.
- Carlson, R. A., Sullivan, M. A., & Schneider, W. (1989). Practice and working memory effects in building procedural skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 517–526.
- Carlson, R. A., & Yaure, R. G. (1990). Practice schedules and the use of component skills in problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 484–496.
- Csikszentmihalyi, M. (1979). Intrinsic rewards and emergent motivation. In M. Lepper & D. Green (Eds.), *The hidden costs of reward: New perspectives on the psychology of human motivation*. New York: Erlbaum.
- Deci, E., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York: Plenum Press.
- Goode, S., & Magill, R. (1986). Contextual interference effects in learning three badminton serves. *Research Quarterly for Exercise and Sport, 53*, 308–314.
- Guadagnoli, M. A., & Lee, T. D. (2004). Challenge point: A framework for conceptualizing the effects of various practice conditions in motor learning. *Journal of Motor Behavior, 36*, 212–224.
- Hebb, D. O. (1955). Drives and the CNS. *Psychological Review, 62*, 243–254.
- Ivanchenko, V., & Jacobs, R. A. (2003). A developmental approach aids motor learning. *Neural Computation, 15*, 2051–2065.
- Kleim, J. A., Barbay, S., Cooper, N. R., Hogg, T. M., Reidel, C. N., Remple, M. S., et al. (2002). Motor learning-dependent synaptogenesis is localized to functionally reorganized motor cortex. *Neurobiology of Learning and Memory, 77*, 63–77.
- Lee, T. D., Magill, R. A., & Weeks, D. J. (1985). Influence of practice schedule on testing schema theory predictions in adults. *Journal of Motor Behavior, 17*, 283–299.
- Liu, Y. T., Mayer-Kress, G., & Newell, K. M. (2003). Beyond curve fitting: A dynamical systems account of exponential learning in a discrete timing task. *Journal of Motor Behavior, 35*, 197–207.
- MacKay, D. J. C. (1992). Information-based objective function for active data selection. *Neural Computation, 4*, 590–604.
- Nelson, T. O., & Leonesio, R. J. (1988). Allocation of self-paced study time and the “labor-in-vain effect.” *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 676–686.
- Nudo, R. J., Wise, B. M., SiFuentes, F., & Milliken, G. W. (1996, June 21). Neural substrates for the effects of rehabilitative training on motor recovery after ischemic infarct. *Science, 272*, 1791–1794.
- Osu, R., Hirai, S., Yoshioka, T., & Kawato, M. (2004). Random presentation enables subjects to adapt to two opposing forces on the hand. *Nature Neuroscience, 7*, 111–112.
- Plautz, E., Milliken, G. W., & Nudo, R. J. (2000). Effects of repetitive motor training on movement representations in adult squirrel monkeys: Role of use versus learning. *Neurobiology of Learning and Memory, 74*, 27–55.
- Pollatou, E., Kioumourtoglou, E., Agelousis, N., & Mavromatis, G. (1997). Contextual interference effects in learning novel motor skills. *Perceptual and Motor Skills, 84*, 487–496.
- Sanger, T. D. (1994). Neural network learning control of robot manipulators using gradually increasing task difficulty. *IEEE Transactions on Robotics and Automation, 10*, 323–333.
- Sanger, T. D. (2004). Failure of motor learning for large initial errors. *Neural Computation, 16*, 1873–1886.
- Schmidt, R. A. (1975). A schema theory of discrete motor skill learning. *Psychological Review, 82*, 225–260.
- Schmidt, R. A., & Lee, T. D. (1999). *Motor control and learning: A behavioral emphasis*. Champaign, IL: Human Kinetics.
- Shapiro, D. C., & Schmidt, R. A. (1982). The schema theory: Recent evidence and developmental implications. In J. A. S. Keslo & J. E. Clark (Eds.), *The development of movement control and co-ordination* (pp. 113–150). New York: Wiley.
- Shea, J. B., & Morgan, R. (1979). Contextual interference effects on the acquisition, retention, and transfer of motor skills. *Journal of Experimental Psychology: Human Learning and Memory, 3*, 179–187.
- Simon, D. A., & Bjork, R. A. (2001). Metacognition in motor learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 907–912.
- Simon, D. A., Cullen, J. D., & Lee, T. D. (2002). Win-shift/lose-stay: Contingent switching as an alternative to random practice? *Journal of Sport and Exercise Psychology, 24*, S116.
- Thrun, S. (2000, March 3). Probabilistic algorithms in robotics. *AI Magazine, 21*(4), 93–109.
- Titze, R., Shea, J. B., & Romack, J. (1993). The effect of learner control on the acquisition and retention of a motor task. *Journal of Sport and Exercise Psychology, 13*, S84.
- Vygotsky, L. S. (1978). *Mind in society*. Cambridge, MA: Harvard University Press.
- Welch, G., & Bishop, G. (2004). *An introduction to the Kalman filter* (Tech. Rep. No. TR 95-041). Chapel Hill: University of North Carolina, Department of Computer Science.
- White, R. H. (1959). Motivation reconsidered: The concept of competence. *Psychological Reviews, 66*, 297–333.
- Winstein, C. J. (1991). Knowledge of results and motor learning: Implications for physical therapy. *Physical Therapy, 71*, 140–149.
- Wu, W. F., Magill, R. A., & Foto, J. G. (2005). Allowing learners to choose: Self-regulated practice schedules for learning multiple movement patterns. *Journal of Sport & Exercise Psychology, 27*, S161.
- Zhang, B. T. (1994). Accelerated learning by active example selection. *International Journal of Neural Systems, 5*, 67–75.

Submitted November 16, 2006

Revised April 24, 2007