1 Texas State University–San Marcos, San Marcos, TX, USA Corresponding Author: Amy L. Simmons, Texas State University–San Marcos, School of Music, 601 University Drive, San Marcos, TX 78666-4616 Email: amy.simmons@txstate.edu

Distributed Practice and Procedural Memory Consolidation in Musicians' Skill Learning

Amy L. Simmons[1]

Abstract This research was designed to determine whether musicians' learning is affected by the time intervals interposed between practice sessions. Twenty-nine non-pianist musicians learned a 9-note sequence on a piano keyboard in three practice sessions that were separated by 5 min, 6 hr, or 24 hr. Significant improvements in performance accuracy were observed in Session 2 only in the group whose sessions were separated by 24 hr. There were significant increases in performance speed in Session 2 in all three practice conditions, results which likely were attributable to the inclusion of all Session 1 data in the analysis. Additional significant speed increases were observed in Session 3 only in the groups whose sessions were separated by 6 and 24 hr. These results suggest that sleep-based procedural memory consolidation may enhance performance accuracy in music skill learning, whereas enhancements in performance speed may be attributable to both wake- and sleep-based consolidation processes.

Keywords distributed practice, procedural skill learning, memory consolidation

Most musicians and teachers accept the notion that music learning is advantaged when practice time is distributed among multiple sessions rather than massed in a single session. The positive effects of distributed practice have been observed in numerous domains of motor skill learning (for reviews, see Donovan & Radosevich, 1999; Lee

& Genovese, 1988; Lee & Wishart, 2005). Surprisingly little research in music has examined this phenomenon directly, although nearly all research in music practice cites as primary evidence for this approach the work of Rubin-Rabson (1940), who reported that pianists who distributed practice over time performed a memorized melody more accurately than did pianists who engaged in an equivalent amount of practice in a single session. Only recently have the processes that underlie the effectiveness of distributed practice in skill learning been elucidated. Musicians know that extended rest intervals between practice sessions provide relief from mental and physical fatigue; less commonly understood is that rest intervals allow time for the neurophysical processes of memory consolidation (for an overview, see Walker, 2005). Theories of memory consolidation were first proposed in 1895 by Oerhn (as cited in Eysenck & Frith, 1977), who observed improvements in performance following periods of rest. More recent investigations of human memory have revealed a great deal about the time course of memory encoding and storage. Although the biological processes of memory consolidation have yet to be characterized fully, it is well understood that acquiring and forming memories for motor skills involve structural and functional reorganization in the brain (Walker & Stickgold, 2006). Consolidation is the mechanism through which motor skills and other memories are encoded and refined, resulting in their resistance to interference and forgetting (Walker, 2005; Walker & Stickgold, 2004). The processes of consolidation are thought to begin during physical practice and to continue covertly after practice ends (Luft & Buitrago, 2005). Wake-based consolidation allows for the intact encoding and storage of fragile new memories. Studies of motor learning in humans have shown that learners who recall newly acquired skills prior to sleep perform with the same levels of accuracy and speed that were achieved by the end of previous practice (Fischer, Hallschmid, Elsner, & Born, 2002; Robertson, Pascual-Leone, & Miall, 2004; Robertson, PascualLeone, & Press, 2004; Shea, Lai, Black, &

Park, 2000; Walker, Brakefield, Hobson, & Stickgold, 2003; Walker, Brakefield, Morgan, Hobson, & Stickgold, 2002; Walker, Brakefield, Seidman, et al., 2003). Memory consolidation processes are further activated during sleep. Sleep-based consolidation that occurs during intervals between practice and recall sessions has been shown to enhance motor skill memory significantly (Cash, 2009; Duke & Davis, 2006; Fischer et al., 2002; Kuriyama, Stickgold, & Walker, 2004; Maquet et al., 2003; Mednick, Nakayama, & Stickgold, 2003; Vertes & Eastman, 2000; Walker, Brakefield, Hobson, et al., 2003; Walker, Brakefield, Seidman, et al., 2003). Research in human movement has demonstrated that learners who distribute practice over time (i.e., dividing practice trials across multiple sessions that span several days) perform better than learners who engage in massed practice (i.e., completing all practice trials in one session on 1 day) when skills are recalled at least 24 hours after practice ends (Dail & Christina, 2004; Shea et al., 2000; Tsutsui, Lee, & Hodges, 1998). In Simmons 359 recent years, researchers in this domain of human learning have attributed the observed performance differences between massed and distributed practice to the processes of memory consolidation (Dail & Christina, 2004; Shea et al., 2000). Many of the studies cited earlier required learners with no skill-related experience to practice relatively simple motor skills (i.e., skills that have one degree of freedom, can be acquired in one practice session, and typically are not skills executed outside laboratories). There is much less research that has examined procedural memory consolidation in the context of more complex motor skills (i.e., skills that include multiple degrees of freedom, require multiple practice sessions for skill acquisition, and are often skills people perform outside of laboratories), a category under which music performance inarguably falls (Wulf & Shea, 2002). Music learners negotiate many skill components simultaneously (e.g., reading notation, translating notation into motor output, processing resulting auditory information, modifying motor behavior based on auditory feedback), and the acquisition of music skills often requires multiple practice sessions. In new learning situations, musicians' perceptions of skill complexity are modified by individual performance ability and prior training and are more aptly characterized as existing on a continuum of skill complexity than by dichotomous terms (e.g., accomplished musicians who are in the early stages of learning a secondary instrument often perceive they are negotiating quite complex motor skills even when performing simple melodies). Evidence for consolidation-based enhancement of complex motor skills has been observed in music learning (Allen, 2007; Duke, Allen, Cash, & Simmons, 2009; Simmons & Duke, 2006). In these studies, nonpianist musicians (who had a moderate amount of secondary-instrument training in piano performance) practiced an unfamiliar keyboard melody in a 12-min training session, then completed a brief retest session following an interval of rest. Simmons and Duke (2006) observed that musicians who slept during the 12-hr rest interval between sessions performed more accurately at retest than did those who remained awake between sessions. Allen (2007) and Duke et al. (2009) reported similar enhancements in the performance of musicians who slept during a 12-hr rest interval between practice and retest. These studies were the first to demonstrate consolidation-based enhancements with experienced learners performing a music skill. The purpose of the present study was to examine musicians' learning of a complex motor skill over multiple equivalent practice sessions separated by different rest interval conditions. There are currently no published studies that compare experienced learners' performance under practice schedules that are either massed (including 5-min rest intervals that do not allow sufficient time for consolidation to observably affect performance; see Muellbacher et al., 2002; Walker, Brakefield, Seidman, et al., 2003) or distributed to include different time intervals of memory consolidation (6-hr and 24-hr rest intervals). I sought to answer the question: Will differences exist in the accuracy and speed of experienced learners'

performances across multiple practice sessions separated by these rest intervals? Method Participants were 291 music majors at The University of Texas at Austin (17 males). All were right-handed, between 18 and 40 years of age, and had no neurological, psychiatric, or sleep-disorder histories. All reported no extensive training or experience on the piano beyond a maximum of five semesters of undergraduate group piano instruction; in those classes, students learned to read, perform, and improvise beginner-level keyboard music and typically demonstrated various levels of rudimentary skill. Graduate and undergraduate music majors were solicited via e-mail and classroom announcement, which provided them with the participation criteria described above and information regarding monetary compensation. Students who fit the criteria and chose to participate gave informed consent. Participants learned a 9-note sequence (see Figure 1) on a digital piano with their left (non-dominant) hand, with the goal of performing the sequence "as quickly, accurately, and evenly as possible." They practiced the sequence in three 15- to 20-min individual sessions. Each practice session consisted of three blocks of 15 performance trials, with each block separated by 30 sec of rest. Each performance trial was followed by 3 sec of silence and an audiovisual cue for the next performance trial to begin. I created three experimental conditions to assess possible effects of sleep- and wake-based memory consolidation on performance accuracy and speed by assigning different rest intervals between sessions: 5 min (massed practice, rest intervals included physical and mental rest, n = 9), 6 hr (distributed practice, rest intervals included wake-based consolidation, n = 10), and 24 hr (distributed practice, rest intervals included both wake- and sleep-based consolidation, n = 10). Participants in the 5-min group completed all three sessions in 1 hour between 4:00 and 8:00 p.m.; sessions for the 24-hr group were scheduled between 10:00 a.m. and 12:00 p.m. on 3 consecutive days. The 6-hr group completed all sessions in 1 day with one session scheduled during each of the following intervals: 8:00-10:00 a.m., 2:00-4:00 p.m., and 8:00-10:00 p.m. Conditions were assigned randomly to participants. Participants agreed to abstain from engaging in behaviors that are known to affect cognitive function and motor performance; specifically, participants agreed to avoid drinking alcoholic and caffeinated beverages and to avoid using other mind-altering drugs for 12 hr prior to and for the duration of their participation in the study. Participants Figure 1.Target melody. Numbers below the notation indicate left hand fingerings; 1 = thumb.

 whose practice sessions were separated by 6 hr agreed to avoid napping between sessions, and participants in the 24-hr group agreed to sleep at night between sessions. I used a Musical Instrument Digital Interface (MIDI) device to connect a Macintosh PowerBook computer to a Roland digital piano for data collection. Notation for the sequence was presented to participants on the laptop computer, which was placed on the keyboard's music rack. Participants listened to all electronic cues (the sound of the piano during every performance trial and a bell-like tone that served as the auditory start cue) through Bose noise-cancelling headphones. The test proctor listened through a second set of headphones. A computer program written specifically for this research using Max/MSP software was set up to display the sequence, to run the protocol, and to collect MIDI performance data. The computer continuously displayed the sequence in music notation for all groups. Also displayed were dots that appeared above each note on the staff. The dots illuminated in red from left to right with each keypress to help participants keep track of where they were in the sequence as they practiced. The lights above each note were illuminated regardless of whether the correct pitch was performed. The computer stopped recording on each trial after the ninth note was played. The Max/MSP software recorded MIDI data during all three sessions. Data for the following variables were analyzed: accuracy, defined as the number of

keypress errors per sequence, and speed, defined as the time elapsed between first and last keypresses in each sequence (expressed in milliseconds). Before participants began each practice session, they rated their feeling of alertness using the Stanford Sleepiness Scale (Hoddes, Zarcone, Smythe, Phillips, & Dement, 1973), answered questions about their music backgrounds (e.g., duration of principal instrument study, secondary instrument training), described how they felt at the moment, and reported on their compliance with study criteria (e.g., no consumption of caffeine, alcohol, drugs), instrumental practice they had engaged in, and sleep time for the previous night. Participants then played through the sequence one time as slowly as needed to play the correct notes with the correct fingerings. If participants struggled with this, feedback was offered and repetitions were allowed until one correct performance of the sequence was achieved. During this time, participants were free to ask questions about the procedure and the sequence. Participants were told, "Do your best to play the melody from beginning to end each time without stopping. Do not practice specific parts out of context or vary the rhythm pattern; in other words, play it just as written." Before data collection began, participants played two test trials with the computer program so they could orient themselves to the way the program worked. They were given another opportunity to ask questions and then began their performance. At the end of Session 1, participants in the 5-min group were given a break, during which time participants were engaged in conversation to prohibit mental and physical practice. Before being dismissed, participants in the 6-hr and 24-hr groups were reminded to abstain from drinking caffeinated or alcoholic beverages and from using other mind-altering substances and to refrain from practicing the sequence between Downloaded from jrm.sagepub.com at PENNSYLVANIA STATE UNIV on May 16, 2016 362 Journal of Research in Music Education 59(4) sessions. Participants in the 6-hr group also were reminded to avoid napping, and those in the 24-hr group were asked to note how long they slept at night. Sessions 2 and 3 were conducted in the same manner as Session 1. Results I found no significant correlations between reported sleep and note accuracy (.05 was the alpha level for all statistical comparisons): Session 1, $r = -.186$, $p = .352$; Session 2, $r = -.582$, $p = .078$; Session 3, $r = .119$, $p = .743$. Likewise, there were also no significant correlations between reports of sleep and speed: Session 1, $r = -.372$, $p = .056$; Session 2, $r = -.217$, $p = .548$; Session 3, $r = -.291$, $p = .414$. I compared participants' reports of alertness on the Stanford Sleepiness Scale given at the beginning of each session with corresponding note accuracy and speed data. There were no significant correlations between reports of alertness and note accuracy: Session 1, $r = -.104$, $p = .590$; Session 2, $r = -.195$, $p = .411$; Session 3, $r = .064$, $p = .790$. Similar results were obtained between reports of alertness and speed: Session 1, $r = -.266$, $p = .163$; Session 2, $r = -.337$, $p = .146$; Session 3, $r = -.274$, $p = .242$. Although there was considerable variation among participants' reports of sleep and alertness, those differences had no consistent effect on performance. Note accuracy. I compared the note accuracy in participants' performances using one-way repeated measures analyses of variance (ANOVAs) for each rest interval condition. In this study, as in Shea et al. (2000), all trials in each block were included in the analyses. Means for each group are presented in Figure 2. I observed a significant difference among the three session means for the 24-hr group, $F(2, 18) = 4.924$, $p = .020$, $\eta^2 p = .354$. Post hoc analyses2 indicate that the mean for the first practice session was significantly higher than the means for Session 2, $p = .033$, and Session 3, $p = .045$, which were not significantly different from one another, $p = .500$. There were no significant differences among the three session means for the 6-hr group, $F(2, 18) = 0.286$, $p = .754$; likewise, there were no significant differences among the three session means for the 5-min group, $F(2, 16) = 0.956$, $p = .405$. It should be noted that the error rate in the 5-min group in the first practice session was near zero, much lower than that of the other two groups. The reason for this difference in the initial block is unknown, but it obviates meaningful comparisons between the accuracy scores of the 5-min group and the

other two groups. Mean note accuracy data for the 6-hr and 24-hr groups, whose Session 1 performances were similar, suggest differences between the effects of wake- and sleep-based consolidation. Put simply, I found significant enhancements in Session 2 performance for participants who slept between sessions. The 24-hr group's data indicate that participants performed with a mean of 10.8 errors in Session 1 (.24 errors per sequence in a 45-trial session); following sleep, these participants performed with a mean of 2.2 errors in Session 2 (.05 errors per sequence). Simmons 363 No group showed evidence of improvements in accuracy between Sessions 2 and 3. It seems that a second night of sleep-based consolidation did not yield detectable performance enhancements in accuracy between Sessions 2 and 3 in the 24-hr group, nor did wake-based consolidation yield Session 3 enhancements in the 6-hr group. Speed. Means for each group are presented in Figure 3. I compared the speed of participants' performances using one-way repeated measures ANOVAs for each rest interval condition and found a significant difference among the three session means in the 5-min group, $F(1, 8)3 = 27.683$, $p = .001$, $\eta2 p = .776$. Post hoc analyses4 indicate that the mean for Session 1 was significantly higher (i.e., participants performed more slowly) than the means for Session 2, $p = .002$, and Session 3, $p < .001$, which were not significantly different from one another, $p = .232$. Similarly, there were significant differences among the three session means in the 6-hr and 24-hr groups: 6-hr group, $F(1, 9)3 = 28.321$, $p < .001$, $\eta2 p = .759$; 24-hr group, $F(1, 9)3 = 18.396$, $p = .002$, $\eta2 p = .671$. Post hoc analyses indicate that the means for Sessions 1, 2, and 3 were all significantly different from one another: 6-hr group, Session 1 versus Session 2, $p < .001$; Session 1 versus Session 3, $p < .001$; and Session 2 versus Session 3, $p = .003$; 24-hr group, Session 1 versus Session 2, $p = .004$; Session 1 versus Session 3, $p = .003$; and Session 2 versus Session 3, $p = .004$. Session means for speed reveal that the extent of performance improvements between Sessions 1 and 2 were similar in all three groups (speed means in Session 2 were 23–26 beats per minute [bpm] faster than Session 1 means), whereas the 6-hr and 24-hr groups showed greater speed gains in Session 3 than did the 5-min group (6-hr 0 0.05 0.1 0.15 0.2 0.25 0.3 0.35 5 min 6 hr 24 hr Mean Errors Per Sequence Group Session 1 Session 2 Session 3 $p = .045$ $p = .033$ Figure 2. Mean errors per sequence for each session. Error bars represent ±1 standard error of the mean. 364 Journal of Research in Music Education 59(4) and 24-hr groups, mean speed increase of 12 bpm; 5-min group, mean increase of only 2 bpm). The increases in speed between Sessions 2 and 3 were expected in the 24-hr condition, based on the results of previous research done with simple motor skills. Session 3 speed enhancements evident in the performances of the 6-hr group are inconsistent with results of studies using non-music tasks but are consistent with the Simmons and Duke (2006) observation that speed enhancements in music learning may be attributable to both wake- and sleep-based consolidation. Discussion My results suggest that performance accuracy is enhanced by sleep-based memory consolidation, a finding that is in line with research conducted using simpler motor skills (Cash, 2009; Duke & Davis, 2006; Fischer et al., 2002; Walker, Brakefield, Hobson, et al., 2003; Walker et al., 2002; Walker, Brakefield, Seidman, et al., 2003) and with research using music tasks similar to the one used in this study (Allen, 2007; Duke et al., 2009; Simmons & Duke, 2006). The effect of memory consolidation on performance speed remains less attributable to sleep-based consolidation, because all three groups demonstrated enhanced performance in Session 2, and both wake- and sleep-based consolidation groups demonstrated speed enhancements in Session 3. The interpretation of the effects of consolidation in the current investigation are somewhat limited by participants having reached very high levels of accuracy early on in practice. In the first session of the 5-min condition and in the second session of the 0 500 1,000 1,500 2,000 2,500 3,000 3,500 4,000 5

min 6 hr 24 hr Mean Sequence Duration (ms) Group Session 1 Session 2 Session 3 p = .002 p < .001 p = .003 p = .004 p = .004 Figure 3. Mean sequence duration for each session. Error bars represent ±1 standard error of the mean.

Simmons 365 24-hr condition, performance accuracy reached an apparent ceiling (approximately 2.25 errors per session), making it unlikely that continued improvements in performance accuracy would be observable in subsequent sessions. Based on the performance accuracy data from Session 1, it appears that, despite the random assignment of conditions, the ability level of the participants in the 5-min group was greater than that of the participants in the other two groups. There are two important differences between the practice procedures of this study and those used in many previous investigations of memory consolidation. First, I controlled the number of repetitions performed in each block of practice. Second, I compared performance across three practice sessions of equal duration and content, a design based on the procedures used by Shea et al. (2000). These data nevertheless suggest effects of memory consolidation similar to reports of other music and non-music investigations that controlled for practice time (not number of trials) and compared performance at the end of practice with performance during a brief retest. I noticed that participants in all groups demonstrated a clear preference for establishing and maintaining note accuracy over increasing performance speed, which is consistent with the reports of Simmons and Duke (2006) and Allen (2007). In informal conversations with me at the end of Session 3, many participants commented that they limited increases in speed for the sake of accuracy. This tendency may be unique to musicians, as it is not typically reported in motor learning research in other domains. This may explain the finding that participants in all conditions made increases in speed between Session 1 (which included the slowest performance trials in the initial stages of acquisition) and Session 2. This is the first study that examines changes occurring in experienced learners' performance across multiple practice sessions that were spaced by very brief intervals of rest (massed practice) or by extended rest intervals (distributed practice) that allowed time for wake-based consolidation only or for both wake- and sleep-based consolidation. The findings from this and other investigations, which suggest that sleep-based consolidation may enhance performance accuracy and that both sleep- and wake-based consolidation may enhance speed, contribute to our understanding of human cognition, skill development, and memory formation. Explanations of the neurophysical processes that encode, refine, and retrieve memories are beginning to clarify the bases for the observation that distributing practice across time is an advantageous course of action in nearly all domains of human learning. The effect of memory consolidation on complex motor skill learning has potential implications for instrumental music study, as these findings inform discussions of practice organization and scheduling. The efficiency of distributed practice may increase levels of students' enjoyment and motivation and also may provide students with more musically gratifying practice experiences and a greater sense of accomplishment than might otherwise be perceived with less effective practice strategies. Understanding the mechanisms by which distributed practice benefits learning not only leads to more informed planning and decision making by teachers but also may lead to more interesting and rewarding practice experiences for learners. 366 Journal of Research in Music Education 59(4) Studies of the most basic processes of procedural memory formation may prompt future explorations into memory processes that underlie more complex music learning. Results like those reported here and in related investigations undoubtedly will lead to additional research that will address how music teachers and performers may structure individual and group practice most effectively to exploit the neural processes that underlie skill development. Author's Note This article is based on the author's doctoral dissertation, "Effects of Practice Variability and Distribution

Notes 1. Data were collected from 31 participants. Out of 18 total data points (3 practice block means from each of 3 sessions for 2 dependent measures), the scores of two participants (one from the 5-min group and one from the 6-hr group) were greater than two standard deviations away from the group mean for each data point in at least 5 of the 18 blocks. They were considered outliers and their data were dropped from consideration. 2. Paired samples t test (one-tailed) with Bonferroni correction. 3. Corrected df for violation of the sphericity assumption (Greenhouse-Geisser). 4. Paired samples t test (one-tailed) with Bonferroni correction.

References Allen, S. E. (2007). Procedural memory consolidation in musicians (Doctoral dissertation, The University of Texas at Austin). Retrieved from http://search.proquest.com/docview/304828 824? accountid=5683 Cash, C. D. (2009). Effects of early and late rest intervals on performance and overnight consolidation of a keyboard sequence. Journal of Research in Music Education, 57, 252–266. doi:10.1177/0022429409343470 Dail, T. K., & Christina, R. W. (2004). Distribution of practice and metacognition in learning and long-term retention of a discrete motor task. Research Quarterly for Exercise and Sport, 75, 148–155. Simmons 367 Donovan, J. J., & Radosevich, D. J. (1999). A meta-analytic review of the distribution of practice effect: Now you see it, now you don't. Journal of Applied Psychology, 84, 795–805. doi:10.1037/0021-9010.84.5.795 Duke, R. A., Allen, S. E., Cash, C. D., & Simmons, A. L. (2009). Effects of early and late breaks during training on overnight memory consolidation of a keyboard melody. Neurosciences and Music III: Disorders and Plasticity. Annals of the New York Academy of Sciences, 1169, 169–172. doi:10.1111/j.1749-6632.2009.04795.x Duke, R. A., & Davis, C. M. (2006). Procedural memory consolidation in the performance of brief keyboard sequences. Journal of Research in Music Education, 54, 111–124. doi:10.1177/002242940605400203 Eysenck, H. J., & Frith, C. D. (1977). Reminiscence, motivation, and personality: A case study in experimental psychology (Vol. 21). Oxford, England: Plenum. Fischer, S., Hallschmid, M., Elsner, A. L., & Born, J. (2002). Sleep forms memory for finger skills. Proceedings of the National Academy of Sciences, 99, 11987–11991. doi:10.1073/ pnas.182178199 Hoddes, E., Zarcone, V., Smythe, H., Phillips, R., & Dement, W. C. (1973). Quantification of sleepiness: A new approach. Psychophysiology, 10, 431–436. doi:10.1111/j.1469-8986.1973 .tb00801.x Kuriyama, K., Stickgold, R., & Walker, M. P. (2004). Sleep-dependent learning and motor-skill complexity. Learning & Memory, 11, 705–713. doi:10.1101/lm.76304 Lee, T. D., & Genovese, E. D. (1988). Distribution of practice in motor skill acquisition: Learning and performance effects reconsidered. Research Quarterly for Exercise and Sport, 59, 277–287. Lee, T. D., & Wishart, L. R. (2005). Motor learning conundrums (and possible solutions). Quest, 57, 67–78. Luft, A. R., & Buitrago, M. M. (2005). Stages of motor skill learning. Molecular Neurobiology, 32, 205–216. doi:10.1385/MN:32:3:205 Maquet, P., Laureys, S., Perrin, F., Ruby, P., Melchior, G., Boly, M., . . . Peigneux, P. (2003). Festina lente: Evidences for fast and slow learning processes and a role for sleep in human motor skill learning. Learning & Memory, 10, 237–239. doi:10.1101/lm.64303 Mednick, S. C., Nakayama, K., & Stickgold, R. (2003). Sleep-dependent learning: A nap is as good as a night. Nature Neuroscience, 6, 697. doi:10.1038/ nn1078 Muellbacher, W., Ziemann, U., Wissel, J., Dang, N., Kofler, M., Facchini, S., . . . Hallett, M. (2002). Early consolidation in human primary motor cortex. Nature, 415, 640–

644. doi:10.1038/ nature712 Robertson, E. M., Pascual-Leone, A., & Miall, R. C. (2004). Current concepts in procedural consolidation. Nature Reviews Neuroscience, 5, 576–582. doi:10.1038/nrn1426 Robertson, E. M., Pascual-Leone, A., & Press, D. Z. (2004). Awareness modifies the skill-learning benefits of sleep. Current Biology, 14, 208–212. doi:10.1016/ j.cub.2004.01.027 Rubin-Rabson, G. (1940). Studies in the psychology of memorizing piano music, II: A comparison of massed and distributed practice. Journal of Educational Psychology, 31, 270–284. doi:10.1037/h0061174 Downloaded from jrm.sagepub.com at PENNSYLVANIA STATE UNIV on May 16, 2016 368 Journal of Research in Music Education 59(4) Shea, C. H., Lai, Q., Black, C., & Park, J. H. (2000). Spacing practice sessions across days benefits the learning of motor skills. Human Movement Science, 19, 737–760. doi:10.1016/ S0167-9457(00)00021-X Simmons, A. L., & Duke, R. A. (2006). Effects of sleep on performance of a keyboard melody. Journal of Research in Music Education, 54, 257–269. doi:10.1177/002242940605400308 Tsutsui, S., Lee, T. D., & Hodges, N. J. (1998). Contextual inference in learning new patterns of bimanual coordination. Journal of Motor Behavior, 30, 151–157. doi:10.1080/ 00222899809601332 Vertes, R. P., & Eastman, K. E. (2000). REM sleep is not committed to memory. Behavioral & Brain Sciences, 23, 1057–1063. doi:10.1017/S0140525X00984024 Walker, M. P. (2005). A refined model of sleep and the time course of memory formation. Behavioral and Brain Sciences, 28, 51–104. doi:10.1017/S0140525X05000026 Walker, M. P., Brakefield, T., Hobson, J. A., & Stickgold, R. (2003). Dissociable stages of human memory consolidation and reconsolidation. Nature, 425, 616–620. doi:10.1038/nature01930 Walker, M. P., Brakefield, T., Morgan, A., Hobson, J. A., & Stickgold, R. (2002). Practice with sleep makes perfect: Sleep-dependent motor skill learning. Neuron, 35, 205–211. doi:10.1016/ S0896-6273(02)00746-8 Walker, M. P., Brakefield, T., Seidman, J., Morgan, A., Hobson, J. A., & Stickgold, R. (2003). Sleep and the time course of motor skill learning. Learning & Memory, 10, 275–284. doi:10.1101/ lm.58503 Walker, M. P., & Stickgold, R. (2004). Sleep-dependent learning and memory consolidation. Neuron, 44, 121–133. doi:10.1016/ j.neuron.2004.08.031 Walker, M. P., & Stickgold, R. (2006). Sleep, memory, and plasticity. Annual Review of Psychology, 57, 139–166. doi:10.1146/annurev.psych.56.091103.070307 Wulf, G., & Shea, C. H. (2002). Principles derived from the study of simple skills do not generalize to complex skill learning. Psychonomic Bulletin & Review, 9, 185–211. doi:10.3758/ BF03196276
Bio Amy L. Simmons is assistant professor of music education at Texas State University. Her research focuses on motor skill acquisition and memory development in the context of music performance and on identifying characteristics of expert teaching.

Abstract

Motor performance induces a postexercise increase in corticomotor excitability that may be associated with motor learning. We investigated whether there are hemispheric differences in the extent and/or time course of changes in corticomotor excitability following a manipulation task (Purdue pegboard) and their relationship with motor performance. Single- and paired-pulse (3 ms) transcranial magnetic stimulation (TMS) was used to assess task-induced facilitation of the muscle evoked potential (MEP) and intracortical inhibition (ICI) for three intrinsic hand muscles acting on digits 1, 2, and 5. Fifteen right-handed subjects performed three 30-s pegboard trials with left or right hand in separate sessions. TMS was applied to contralateral motor cortex before and after performance. Number of pegs placed was higher with the right hand, and performance improved (motor learning) with both hands over the three trials. MEP facilitation following performance was short-lasting (<15 min), selective for muscles engaged in gripping the pegs, and of similar magnitude in left and right

hands. ICI was reduced immediately following performance with the right hand, but not the left. The extent of MEP facilitation was positively correlated with motor learning for the right hand only. We conclude that the pegboard task induces a selective, short-lasting change in excitability of corticospinal neurons controlling intrinsic hand muscles engaged in the task. Only left hemisphere changes were related to motor learning. This asymmetry may reflect different behavioral strategies for performance improvement with left and right upper limb in this task or hemispheric differences in the control of skilled hand movements.

INTRODUCTION

Motor practice induces plastic changes in primary motor cortex (M1) which can be studied with transcranial magnetic stimulation (TMS). These changes include an expansion of motor representations (Karni et al. 1995; Nudo et al. 1996; Pascual-Leone et al. 1994, 1995), changes in the kinematics of movements evoked by TMS (Bütefisch et al. 2000; Classen et al. 1998), and facilitation of the muscle evoked potential (MEP) in task muscles tested at rest for ≤60 min after the practice session (Bütefisch et al. 2000; Caramia et al. 2000; Classen et al. 1998; Hauptmann et al. 1997; Muellbacher et al. 2001; Pascual-Leone et al. 1995; Rossi et al. 1999; Samii et al. 1998; Ziemann et al. 2001). The latter phenomenon has been termed postexercise facilitation in some studies, however, we will use the term task-induced MEP facilitation in the present study, as it can be induced by motor imagery of a complex task in the absence of muscle activation (Pascual-Leone et al. 1995) and is not seen following nonfatiguing muscle contraction evoked by peripheral nerve stimulation (Ziemann et al. 2001). There is increasing evidence that task-induced changes in M1 are important for motor learning (Donoghue and Sanes 1994; Karni et al. 1998; Sanes and Donoghue 2000). Several studies have shown that task-induced MEP facilitation is positively correlated with improvements in performance (Pascual-Leone et al. 1995; Muellbacher et al. 2001; Ziemann et al. 2001), and it has recently been suggested that task-induced MEP facilitation may reflect the early consolidation stage of motor skill acquisition in M1 (Muellbacher et al. 2002).

None of the previous studies have addressed whether there are hemispheric differences in the extent or time course of task-

induced MEP facilitation or whether the relationships between the changes in corticomotor excitability and task performance or motor learning differ when using the preferred or nonpreferred hand. The representation of hand muscles and synaptic interconnectivity in primary motor cortex is larger in the dominant (left) hemisphere than the nondominant (right) hemisphere (Hammond 2002). In addition, activation of contralateral M1 is not symmetrical for unimanual tasks performed with left or right hand (Brouwer et al. 2001; Semmler and Nordstrom 1998; Volkmann et al. 1998), so there is reason to suspect that the extent of task-induced MEP facilitation may differ for performance with left and right hands. Another possible source of asymmetry is that the ipsilateral sensorimotor cortex plays a greater role in finger movements of the left hand than the right (Chen et al. 1997; Kim et al. 1993; Ziemann and Hallett 2001). If task-induced MEP facilitation reflects cortical plasticity related to motor learning, as has been proposed, hemispheric differences in these processes may contribute to asymmetries in hand preference for skilled motor tasks. In the present study we have begun to address this question by comparing task-induced MEP facilitation and performance on a fine manipulation task (Purdue pegboard) (Tiffen 1968) for both hands of right-handed subjects. To provide information about the specificity of the changes in each hemisphere, we studied two muscles used to manipulate the pegs, first dorsal interosseous (FDI) and abductor pollicis brevis (APB), and the abductor digiti minimi (ADM) muscle, which was commonly active during the task but did not contribute directly to performance.

Finally, task-induced MEP facilitation is influenced by GABAergic intracortical inhibition (ICI) (Ziemann et al. 2001). Disinhibition may play an important role in motor learning by unmasking excitatory connections to corticospinal neurons in M1 and/or promoting plastic changes at synapses by facilitating long-term potentiation (Hess and Donoghue 1994; Hess et al. 1996; Jones 1993). Hemispheric differences in ICI might

contribute to differences in motor skill between hands, but a relationship between ICI and skilled motor performance of left and right hands has not been assessed in previous studies. To explore the role and specificity of ICI in task-induced MEP facilitation and motor learning with preferred and nonpreferred hands we used a paired-pulse TMS paradigm (Kujirai et al. 1993) to assess ICI for the three muscles before and after pegboard performance with each hand.

METHODS

Fifteen healthy adult volunteers (10 male, 5 female), including two of the authors, ranging in age from 18 to 44 yr gave informed written consent for participation in the experiment. Subjects were right-hand dominant according to the Edinburgh Handedness Inventory (Oldfield 1971) with a median Laterality Quotient (LQ) of 0.80 (range 0.55–1.00). Nine of the 15 subjects had an LQ > 0.80. The experimental procedures conformed with the declaration of Helsinki, and the protocol was approved by the Human Research Ethics Committee of the University of Adelaide.

Apparatus

Subjects were seated in a dental chair with their right or left arm resting comfortably on a Purdue pegboard (Lafayette Instruments model 32020, Lafayette, IN) that was firmly clamped to a table. The height of the chair and position of the pegboard were adjusted to allow unrestricted, comfortable movement of the arm over the entire surface of the pegboard. The surface electromyogram (EMG) was recorded from the left and right FDI, APB, and ADM muscles using bipolar Ag-AgCl electrodes. The signals were amplified (×200–2000), filtered (20 Hz to 1 kHz), and recorded on tape (Vetter model 400D, 22 kHz/ch). Peristimulus epochs of data were digitized (2 kHz/ch sampling rate) and averaged on-line (Labview). TMS was applied using two Magstim 200 (Magstim, Dyfed, UK) stimulators connected via a Bistim module to a single 90-mm

circular coil. This allowed delivery of single- or paired-pulse [3-ms interstimulus interval (ISI)] TMS to assess both test MEP size (single test pulse) and effectiveness of ICI by comparison of size of MEPs in conditioned (paired-pulse) and single-pulse trials (Kujirai et al. 1993). Two oscilloscopes positioned directly in front of the subject displayed EMG from FDI, APB, and ADM of the task hand. Subjects were instructed to monitor these signals to maintain electrical silence in all muscles during delivery of TMS. This was necessary for testing of ICI as it is known to be suppressed during muscle activation (Ridding et al. 1995). Surface electrodes attached over the ulnar and median nerves at the wrist were used with a Digitimer D180 electrical stimulator (Digitimer, UK) to deliver supramaximal stimulation (0.5 Hz, pulse duration 100 μs) of the nerves to elicit maximal M-waves in each muscle.

Protocol

The left and right hands were tested in separate sessions on the same day. The hand tested first was alternated across subjects. Maximal M-waves were recorded for each muscle prior to (Pre) and following (Post) TMS testing of each hand. The optimal location for TMS was identified by initially placing the circular coil at the vertex and then moving it in small steps until the site from which the largest MEPs in all three muscles was found using a moderately suprathreshold intensity. This location was marked on the scalp to allow consistent coil placement during the experiment. The current direction in the coil (viewed from above) was counterclockwise for right-side muscles (left motor cortex) and clockwise for left-side muscles (right motor cortex). These coil orientations induce current flow in the brain that traverses the central sulcus of the hemisphere contralateral to the test muscles in a posterior–anterior direction, which is optimal for production of a MEP at lowest threshold with TMS (Day et al. 1989). Resting motor threshold was defined as the weakest TMS intensity (2% increments) that yielded ≥3 of 5 MEPs > 50 μV in relaxed FDI. Active threshold (3/5 MEPs >100 μV) was assessed with weak voluntary

activation of FDI. Test TMS intensity was adjusted to produce an MEP of approximately 0.5 mV in relaxed FDI, and reliable responses in the other two muscles. ICI was assessed using a paired-pulse paradigm (Kujirai et al. 1993) with a 3-ms ISI. The intensity of the conditioning stimulus in the paired-pulse trials was below active motor threshold and adjusted to produce approximately 50% reduction of the FDI MEP at rest prior to the pegboard task.

Subjects performed three 30-s trials of the Purdue pegboard in which they picked individual pegs from a well using the thumb and index finger and placed them in individual holes in the pegboard. Robust MEP facilitation is induced in FDI following performance of the pegboard task (Rossi et al. 1999). The intertrial interval was approximately 10–20 s. Subjects were encouraged to place as many pegs as possible and the number of pegs placed on each trial was recorded. No practice trials were provided. Prior to and following completion of the pegboard task, subjects were instructed to keep their hand muscles relaxed aided by visual feedback of EMG from the three muscles of interest. EMG from the three muscles was recorded on tape during task performance.

TMS was delivered with all muscles at rest at three times with respect to pegboard performance: prior to the first pegboard trial (Pre), immediately following the final pegboard trial (+0), and 15 min following the final trial (+15). Each block consisted of 24 trials in which single- or paired-pulse TMS (3-ms ISI) was delivered under computer control (randomized), with an intertrial interval of 5 s. Each block, therefore, took 2 min to complete.

After the final block of TMS, M-waves were assessed a second time to determine any peripheral neuromuscular effects of pegboard performance. After a rest period of approximately 10 min, the entire sequence was repeated with the other hand.

Analysis

Between-hemisphere differences for resting motor threshold and test TMS intensity were assessed using paired *t*-test. MEP and M-wave areas were calculated from the averaged rectified EMG. Effectiveness of ICI was quantified as the ratio of the conditioned and test MEP areas, expressed as a percentage. Performance of the pegboard task was quantified as the sum total of pegs placed over the three 30-s trials (Pegboard score). Within-session relative improvement of pegboard performance (Improvement score) was given by the equation (pegs placed in trial 3–pegs placed in trial 1)/pegs placed in trial 1.

Statistical analyses of the effects of Hand (L, R), muscle (FDI, APB, ADM), and Time (Pre, +0, +15) on test MEP size and effectiveness of ICI were performed using repeated-measures ANOVA followed by pairwise posthoc comparisons using Tukey's honestly significant difference multiple comparison procedure. Linear regression was used to analyze the relationship between changes in test MEP size and ICI immediately following the pegboard task (+0) and their relationship with task performance. For these comparisons, MEP data were normalized by using the ratio of the value obtained immediately following the pegboard task (+0) and the baseline period preceding the task (Pre). Data are reported as mean ± SE unless otherwise stated. All statistical tests were considered significant at $P < 0.05$.

RESULTS

Performance measures

Performance in the pegboard task is summarized in Fig. 1*A*. Performance was better overall with the right hand [$F(1,14) = 6.56$, $P = 0.023$]. The mean number of pegs placed by the right hand for the three trials was 42.9 ± 1.4 while the left hand placed 40.7 ± 1.3. Performance improved monotonically across the three trials [$F(2,28) = 40.04$, $P < 0.0001$] with scores on the second and third trials being significantly better than the

preceding trial (Tukey's test; $P < 0.05$; 1st, 12.7 ± 0.4; 2nd, 14.2 ± 0.3; 3rd, 14.9 ± 0.3). The hand by trial interaction was not significant [$F(2,28) = 0.2$, $P = 0.813$], indicating that the improvement in performance was similar for both the left and right hands.

fig. 1.

A: mean pegboard score (±SE) of the left (○) and right hand (•) for each pegboard trial. Performance was significantly better with the right than the left hand (ANOVA, $P < 0.05$). Performance was significantly better on the 2nd and 3rd trials than the preceding trial (ANOVA, $P < 0.05$). *B*: there was a significant negative relationship between pegboard score and the magnitude of improvement ($r = -0.40$, $P < 0.03$) for the pooled data for all subjects. Those subjects who had the highest overall score tended to have the least improvement over the 3 trials.

Download figureDownload PowerPoint

Although there was a significant overall tendency for task performance to improve with successive trials, performance improvement varied over a large range in different subjects. Linear regression revealed a significant negative relationship between the Pegboard Score and Improvement Score from both hands of the 15 subjects (Fig. 1*B*; $r = -0.40$, $P = 0.03$). That is, those subjects with the greatest improvement over the three trials also tended to have the lowest overall score. This relationship was not significant for the left hand ($r = -0.33$, $P > 0.2$), but approached a conventional level of significance for the right hand ($r = -0.49$, $P = 0.06$).

M-waves

M-wave area was analyzed using separate two-way ANOVAs for each muscle with hand (left, right) and time (pre, post) as factors. The time factor was not significant for any muscle ($P > 0.05$) nor was the main effect of hand or the interaction ($P > 0.05$). Thus performance of the pegboard task had no

significant effect on M-wave area.

TMS intensity

Mean resting motor thresholds (expressed as a percentage of maximum stimulator output) were 58 ± 3% for the left hand and 56 ± 2% for the right hand. This difference was not significant ($t_{14} = 1.22$, $P = 0.24$). Test TMS intensities were 70 ± 3% for the left hand and 68 ± 3% for the right hand, a nonsignificant difference ($t_{14} = 0.76$, $P = 0.46$). Relative to resting threshold, test TMS intensities were 121 ± 2% for the left hand and 121 ± 3% for the right hand. This difference was not significant ($t_{14} = 0.18$, $P = 0.86$). Conditioning TMS intensities were 78 ± 2 and 80 ± 2% of active threshold for the left and right hands, respectively. This difference was not significant ($t_{14} = 0.96$, $P = 0.35$). Thus corticomotor excitability and TMS intensities used for the experiments were equivalent for left and right hemispheres.

Changes in MEPs following the pegboard task

Figure 2*A* is an example of EMG activity in the three muscles of the right hand during performance of the pegboard task. Both FDI and APB show robust, phasic activation consistent with their role in peg retrieval and manipulation. By contrast, activity in ADM is substantially weaker and more sustained. Figure 2*B* shows the MEPs obtained in the same three muscles at each of the three time points. Immediately following performance of the pegboard task (+0) there was a clear facilitation of the MEP in both FDI and APB. For both muscles MEP size returned to baseline within 15 min (+15). There was no facilitation of the ADM MEP following the pegboard task.

[?]

fig. 2.
A: electromyogram (EMG) recorded from 1 subject during performance of the pegboard task. Large, phasic bursts of EMG activity were seen in first dorsal interosseous (FDI) and abductor pollicis brevis (APB), with weak EMG activity in abductor digiti minimi (ADM). *B*: averaged muscle evoked potentials (MEPs; test alone) from the muscles of the subject in *A*. Task-induced

MEP facilitation is clearly evident in both FDI and APB immediately following the pegboard task (+0). MEPs in both muscles returned to baseline by 15 min (+15). There is no evidence of MEP facilitation in the ADM muscle at either +0 or +15. Calibration bars amplitude: FDI and APB 500 µV; ADM 125 µV, time = 25 ms.

Figure 3 shows the mean MEP area in the three muscles of the left and right hands at each time point for the 15 subjects. As expected, mean MEP area increased significantly following performance of the pegboard task [$F(2,28) = 6.93$, $P = 0.004$], consistent with the findings of Rossi et al. (1999). MEPs immediately following the pegboard (+0: 5.01 ± 0.48 mV ms) were significantly larger than those obtained prior to performance (Pre: 3.83 ± 0.31 mV ms) and following 15 min rest (+15: 3.79 ± 0.32 mV ms). There was no significant difference between the Pre and +15 MEPs, indicating that task-induced MEP facilitation was relatively short lasting. However, as seen in Fig. 3, the three muscles were not equally affected by the pegboard task, an observation confirmed by a significant muscle by time interaction [$F(4,56 = 3.00$, $P = 0.026$].

fig. 3.
Mean MEP area (±SE) for FDI, APB, and ADM of the left and right hand ($n = 15$). There was significant task-induced MEP facilitation immediately following the pegboard task (+0) in both FDI and APB, but not ADM. MEP area had returned to prepegboard size after 15 min (+15). The magnitude of MEP facilitation was similar in the left and right hands.

Both muscles engaged in manipulation of the pegs, FDI and APB, showed significant task-induced MEP facilitation at +0. MEPs in FDI were 38% larger than prepegboard MEPs, while MEPs in APB increased by 39%. MEPs in both muscles returned to prepegboard size by +15 (change from prepegboard: FDI = –3%, APB = 1%). By contrast, pegboard

performance had no significant effect on the ADM MEP at either +0 (+10%) or +15 (–1%), even though ADM was active during task performance. Thus MEP facilitation induced by pegboard performance appears to be selective for hand muscles engaged in manipulating the pegs and is relatively short lasting (<15 min).

There was no overall difference in the size of MEPs in the left and right hands [$F(1,14) = 0.04$, $P = 0.84$] nor were there any significant between-hand differences for the muscle by time interaction [$F(4,56) = 0.52$, $P = 0.73$].

ICI

The effectiveness of ICI was assessed from the ratio of the MEP size in conditioned and test-alone trials. The effects of pegboard performance on ICI are summarized in Fig. 4. Pegboard performance reduced the effectiveness of ICI [$F(2,28) = 4.03$, $P = 0.029$]. This effect was uniform for the three muscles [$F(4,56) = 1.2$, $P = 0.34$], but differed for left and right hands [$F(2, 28) = 3.42$, $P = 0.047$]. Posthoc tests revealed a significant reduction of ICI effectiveness for the right hand immediately following the pegboard task (+0: 0.66 ± 0.04) compared with the prepegboard trial (Pre: 0.50 ± 0.04). There was a tendency for reduced ICI effectiveness after the pegboard task for both hands at both time points, but the other changes were not significant (Tukey's test; $P > 0.05$).

fig. 4.
Conditioned MEP size (±SE) at Pre, +0, and +15 for the left and right hands. Data are pooled for the 3 muscles. The effectiveness of conditioning TMS was significantly reduced at +0 for the right hand ($P < 0.05$).

Download figureDownload PowerPoint

Relationship between task-induced changes in MEP size and ICI

effectiveness

The relationship between the change in test MEP size and ICI immediately following the pegboard task (+0/Pre) was assessed with linear regression. For all muscles in both hands, these variables tended to be negatively related. However, none of the regressions were statistically significant ($P > 0.05$).

Relationship between task-induced MEP facilitation and task performance

Significant MEP facilitation was observed in FDI and APB immediately following pegboard performance (+0 in Fig. 3). To examine the relationship between pegboard performance and task-induced MEP facilitation we calculated the ratio of the MEP size at +0 and Pre for the three muscles of each subject, and performed linear regression analysis with the total number of pegs placed (Fig. 5) and the within-session improvement in performance (Fig. 6).

fig. 5.

Regression plots of MEP facilitation immediately following the pegboard task (+0 MEP/Pre MEP) versus Pegboard Score. Task-induced MEP facilitation was negatively correlated with Pegboard Score for all 3 muscles of the right hand (•), and this was significant for FDI. Thus subjects who placed the most pegs tended to have the least MEP facilitation. None of the regressions were significant for muscles of the left hand (○).

Download figureDownload PowerPoint

fig. 6.

Regression plots of MEP facilitation immediately following the pegboard task (+0 MEP/Pre MEP) versus Improvement Score. Task-induced MEP facilitation in FDI and APB of the right hand (•) was significantly positively correlated with Improvement Score. None of the regressions were significant for right ADM or any muscles of the left hand (○). Subjects with the largest gains in performance over the 3 trials had the largest MEP facilitation, but only in muscles of the right hand engaged in the task.

Download figureDownload PowerPoint

Figure 5 shows the relationship between Pegboard Score (total number of pegs placed) and task-induced MEP facilitation for each muscle of both hands. A higher level of proficiency on the pegboard task, as indicated by more pegs placed, tended to be associated with low MEP facilitation. This relationship was only significant, however, for right FDI (Fig. 5).

Figure 6 shows the relationship between the within-session improvement in performance (Improvement Score) and task-induced MEP facilitation. There was a significant positive relationship between these variables in FDI and APB, but only for the right hand. Subjects with a large improvement in performance tended to have greater MEP facilitation immediately following the task, but only for muscles of the right hand engaged in manipulation of the pegs. The coefficient of determination (r^2) indicates that the relative performance improvement accounted for 49% of the variance in the right FDI MEP facilitation and 28% for right APB.

The relationship between changes in effectiveness of ICI immediately following the pegboard task (+0/Pre) and performance was examined in the same manner as the Figs. 5 and 6 analyses. There was no significant relationship between the changes in ICI and Pegboard Score or Improvement Score for any hand or muscle.

DISCUSSION

The principal finding of the present study is a hemispheric asymmetry of the relationship between task-induced MEP facilitation following performance of a fine motor task and within-session improvement of performance (motor learning). Performance of the pegboard task with the contralateral hand induced an increase in corticomotor excitability that was similar in magnitude for the left and right hemispheres, however, only

changes in the left M1 were related to performance. There was a strong positive relationship between the change in excitability of FDI and APB representation in the left hemisphere and within-session improvement in performance with the right (preferred) hand. Consistent with earlier findings (Muellbacher et al. 2001), the changes in excitability were specific for the corticomotor representation of hand muscles used for task performance (FDI and APB), as no change was seen for the ADM muscle, which was active during the task but did not contribute to performance. As rapid changes in M1 corticomotor excitability with motor practice has been found to be related to motor learning (Muellbacher et al. 2001, 2002), the present finding points to a different strategy for performance improvement with left and right hands in this task and/or a hemispheric asymmetry in the role of sensorimotor cortex in motor learning of hand skills.

Mechanisms of task-induced MEP facilitation

Practice of a variety of simple motor tasks induces a facilitation of the MEP in task muscles tested with TMS at rest immediately after the practice session, and this effect lasts for ≤60 min. Examples include steady isometric contraction of wrist muscles (Rossi et al. 1999; Samii et al. 1998), ballistic elbow flexion movements (Ziemann et al. 2001), ballistic pinch grip (Muellbacher et al. 2001), and isolated thumb movements (Bütefisch et al. 2000; Classen et al. 1998). Performance of more complex tasks such as the Purdue pegboard (Rossi et al. 1999; present study) or sequential finger tapping (Pascual-Leone et al. 1995) also facilitates hand muscle MEPs. In the latter study, mental imagery without physical practice also induced MEP facilitation immediately after the training session. These facilitatory changes are believed to be cortically mediated, as training has little effect on the MEP produced by brain stem or cervical stimulation (Muellbacher et al. 2001). Further evidence comes from experiments showing that voluntary practice of thumb movements alters the direction of TMS-evoked movements of the thumb, but training in this way

is less effective in altering the direction of thumb movements evoked by transcranial electrical stimulation (TES) (Classen et al. 1998). As TES preferentially activates corticospinal axons directly (Day et al. 1987), the TES results suggest that practice has few aftereffects on segmental excitability. In the present study there was no change in M-waves following the pegboard task, so the MEP facilitation was not due to a change in the peripheral neuromuscular apparatus. The selectivity of task-induced MEP facilitation, which was restricted to hand muscles used to grip the pegs (FDI and APB), excludes nonspecific processes such as arousal or attention, as these would also be expected to affect ADM as well. Active engagement of the motor cortex is necessary for task-induced MEP facilitation—it could not be induced by peripheral nerve stimulation or passive movement alone (Ziemann et al. 2001). While we did not directly assess segmental motor excitability with TES or H-reflexes, previous studies suggest that the changes in corticomotor excitability we observed reflect changes at the level of the cortex.

Task-induced MEP facilitation and motor learning

Currently, the mechanisms responsible for usage-dependent plasticity in the motor cortex are not well understood. However, the data are consistent with a usage-dependent modification of cortical synapses that has been proposed to underlie motor learning (Asanuma and Pavlides 1997; Sanes and Donoghue 2000). Synaptic plasticity in the cortex is influenced by N-methyl-d-aspartate (NMDA) receptor activation (Castro-Alamancos and Connors 1996) and GABAergic disinhibition (Hess and Donoghue 1994; Hess et al. 1996), and practice-induced changes in MEPs and movement kinematics in humans are blocked by dextromethorphan (an NMDA receptor blocker) and reduced by lorazepam (a GABA$_A$ agonist) (Bütefisch et al. 2000).

Motor learning can be defined as the acquisition of new patterns of muscle activation in time and space to improve

performance of a motor task (Sanes and Donoghue 2000). Several TMS studies have demonstrated an association between task-induced MEP facilitation and improvements in performance (Muellbacher et al. 2001; Ziemann et al. 2001). It has recently been suggested that practice-induced changes in M1 excitability reflect the early stages of memory consolidation in M1 of the practiced movement related to motor learning (Muellbacher et al. 2002). This is supported by the finding that repetitive TMS (rTMS) of M1 immediately following motor practice abolishes gains in performance. The disruption of performance gains is not observed when rTMS is delivered following a 6-h consolidation period nor when it is given over cortical regions outside M1 (e.g., parietal cortex) (Muellbacher et al. 2002). Also consistent with this view is the finding that practice-dependent increases in M1 excitability are no longer present once a motor skill has been overlearned (Muellbacher et al. 2001).

Hemispheric differences in task-induced MEP facilitation and its relation with motor learning

In the present study subjects improved their performance of the pegboard task progressively over three trials (Fig. 1). That is, motor learning occurred over trials 1–3, and performance improvement had not plateaued by trial 3. The improvement in performance was similar for left and right hands, although the strategy may have differed (see the first possible explanation below). Overall performance was better with the right hand, consistent with previous findings in right-handed subjects (Brouwer et al. 2001; Tiffen 1968).

The present study is the first to address the issue of whether task-induced plasticity of M1 differs in the left and right hemisphere and more specifically whether the relationships with motor learning show a hemispheric asymmetry. Task-induced MEP facilitation was similar in FDI and APB muscles of right and left hands and was not seen in ADM, which was active during the task but was not critical for task performance

(Fig. 3). At first glance these results suggest no hemispheric asymmetry of task-induced MEP facilitation in the pegboard task and a similar relationship with motor learning since performance improvements were similar with each hand. However, regression analysis of data from individual subjects revealed a strong positive relationship between task-induced MEP facilitation in FDI and APB and performance improvement for the right hand only (Fig. 5). We consider several possible explanations for this difference between left and right hands.

First, the performance improvement of the two hands may be mediated by different behavioral strategies. The pegboard task requires coordinated control of the muscles of the entire arm, not just the index finger and thumb. A complete cycle of peg retrieval and placement consists of several phases, each differentially dependent on the behavior of arm or hand muscles. Retrieving and placing pegs is heavily reliant on manipulative control of the digits, while transport of the hand is governed largely by proximal muscles of the arm. Improvement in pegboard performance could be brought about through modifications of any, or all, of these phases.

As the peg retrieval and placement phases are both time consuming and critical to skilled performance of the pegboard task, it is likely that at least some of the performance improvement of both hands involved modifications of control of the digits in these phases. This could account for task-induced MEP facilitation observed in FDI and APB of both hands. However, modifications of the reaching phase are likely to be asymmetrical. Behavioral studies consistently report a right arm advantage for the control of intersegmental limb dynamics (Bagesteiro and Sainburg 2002; Goodale 1988; Sainburg 2002) and a relative disadvantage for the left arm when "homing in" on the target (Roy et al. 1994). Thus the left arm is likely to complete aiming movements more slowly due to a relative difficulty in accurately terminating the movement. On the basis of these findings it seems possible that a greater proportion of

left arm improvement could have come about through adjustments of the reaching component, which would ultimately mask the relationship between task-induced MEP facilitation in intrinsic hand muscles and improvement in the pegboard score. As we did not monitor the kinematics of pegboard performance in this study, or EMG of more proximal muscles, this possibility will need to be addressed in future investigations.

A second possible explanation for the asymmetry of the relationship between task-induced MEP facilitation and improvement in performance relates to the hemispheric control of goal-directed movements of the right and left hand. A substantial body of evidence indicates that, for right handed subjects, the left sensorimotor cortex contributes to the performance of *both* hands, whereas the right hemisphere is important for use of the left hand, but not the right. Performance of right hand movements is accompanied by activation primarily of the left (contralateral) sensorimotor cortex, whereas performance of left hand movements is associated with activation of both the right (contralateral) and left (ipsilateral) sensorimotor areas (Kawashima et al. 1993). Left M1 is involved in motor imagery of unilateral movements of either hand, whereas right M1 is only involved in imagery of left hand movement (Fadiga et al. 1999). Studies of stroke patients indicate this involvement of the left hemisphere is functionally important. Wyke (1968, 1971) found that unilateral stroke of the right hemisphere was accompanied by a performance deficit only of the contralateral left hand. By contrast, stroke affecting the left hemisphere was associated with performance deficits of *both* hands. Similarly, ideomotor apraxia, which is characterized by disturbances in the performance of goal-directed movements, results from damage to the left frontal and parietal areas (Haaland et al. 2000) and rarely from damage to the right hemisphere. In most people, inactivation of the left hemisphere with intracarotid injection of barbiturate (Wada test) impairs skilled movements of both hands, whereas right-sided injection impairs only left hand movements (Heilman 2000).

Further evidence of the importance of the left M1 in left hand control is provided by TMS studies. Chen et al. (1997), using repetitive TMS, showed that stimulation of the left M1 during a key pressing task disrupted performance with the contralateral right hand as well as the ipsilateral left hand. Stimulation of the right M1 had a much weaker effect on performance with the right hand. Left M1 involvement in left hand movements increased with increasing complexity of the sequence of key presses.

For the pegboard task it seems reasonable to assume that skilled performance of the right hand was governed largely by the left (contralateral) hemisphere. Task-induced facilitation of the MEP in the right hand, reflecting plasticity in left M1, was strongly associated with motor learning. However, this relationship will not be as straightforward following left hand performance since at least some of the neural processes mediating skilled performance with the left hand are present in the left (ipsilateral) M1. The absence of a relationship between task-induced MEP facilitation and performance improvement following left hand use (Fig. 5) may be due to a contribution of the left hemisphere to motor learning of the left hand, which was not assessed with our protocol as TMS was applied only to the hemisphere contralateral to the task hand. This hypothesis could be addressed in future studies by testing corticomotor excitability of both hemispheres with TMS following unimanual performance.

Hemispheric differences in ICI and its relation with motor learning

It has recently been demonstrated that practice-dependent plasticity in M1 is influenced by GABAergic inhibition (Ziemann et al. 2001). MEP facilitation induced in biceps by practice of ballistic elbow flexion movements was increased in the presence of ischemic nerve block of the hand, which is known to suppress GABAergic inhibition (Ziemann et al. 1998). This was accompanied by a concomitant *decrease* in the effectiveness of ICI circuitry assessed with paired-pulse TMS

(Ziemann et al. 2001). By contrast task-induced MEP facilitation was completely blocked by the administration of lorazepam, a GABA$_A$ agonist, with an accompanying *increase* in ICI effectiveness. Thus the effect of practice on M1 excitability can be influenced by altering the effectiveness of ICI circuits. Reduced intracortical inhibition could contribute to task-induced MEP facilitation by unmasking existing excitatory connections to corticospinal neurons that are activated by TMS and/or by enhancing plastic changes in cortical synaptic connections during training by mechanisms such as long-term potentiation (Hess and Donoghue 1994; Hess et al. 1996; Jones 1993).

Ziemann et al. (2001) found that significant MEP facilitation was induced in left biceps by motor practice alone without any change in ICI effectiveness in right M1 evident following task performance. Similar results were observed for left hand muscles in the present study. Task-induced MEP facilitation is therefore not dependent on a *concurrent* GABAergic disinhibition and unmasking of latent excitatory synaptic connections in M1. ICI effects are suppressed during voluntary activation (Ridding et al. 1995), however, and this may help to fractionate hand muscle activity (Zoghi et al. 2003) as well as promote plastic changes in M1 that persist beyond task performance once ICI returns to baseline levels.

ICI does not differ between hemispheres at rest (Cicinelli et al. 2000; Civardi et al. 2000; Maeda et al. 2002), but this has not been assessed during or after voluntary activation of the muscles. When the right hand was used to perform the pegboard task, ICI effectiveness in left M1 was significantly reduced in the period immediately following the task (Fig. 4). No change in ICI effectiveness was observed in the right hemisphere, however, following left hand performance. Unlike the task-induced MEP facilitation, the task-induced suppression of ICI was not selective for the corticospinal neurons controlling FDI and APB muscles (i.e., the muscle by hand interaction was not significant) nor was its magnitude associated with task

performance. The size of task-induced MEP facilitation was not related to the amount of suppression of ICI following pegboard performance. Our results indicate that motor practice is more effective in suppressing ICI circuits in left M1 than right M1 in the immediate postpractice period. At present the functional significance of this hemispheric difference in ICI for plasticity of corticomotor representations in M1 and motor learning remains unclear.

In summary, brief performance of a fine-motor manipulation task induces transient increases in excitability of the contralateral M1 regardless of whether the task is performed with the left or right hand. The task-induced changes observed with single TMS are similar in magnitude, duration, and selectivity in both hemispheres, although only changes in the left hemisphere were directly correlated with the within-session improvement in performance (motor learning). Intracortical inhibition was suppressed immediately following the task, but only in the left M1 and not selectively for the cortical representation of the task muscles. These asymmetries may reflect different behavioral strategies involving arm and hand muscles that mediate performance improvement with left and right upper limb or hemispheric differences in the control of skilled hand movements. Future studies in this area will need to take these asymmetries into account.

We investigated the acquisition of music performance skills in novice and expert pianists. Temporal disruptions in novice performances coincided with constraints in planning capacities. Child and adult pianists ranging in age (9±26 years), training (3±15 years) and sightreading ability learned to perform a novel musical piece in eleven practice trials. Computerdetected pitch and timing errors revealed: (1) gradual improvements in performance tempo and pitch accuracy with skill level and practice, generally ®tting a power function; (2) a relative-timing/pitch accuracy trade-off and high incidence of simultaneous pitch/time errors; (3) improvements in relative timing (temporal continuity, underlying beat, metrical structure) with skill and practice; and (4) increased anticipatory behavior and a greater range of planning with skill and practice. A strong positive relationship between the mastery of temporal constraints and planning

abilities within performance suggests that these two cognitive indicators are closely related and may arise from segmentation processes during performance. Examination of sequence timing may explicate planning abilities that underlie many complex skills. Keywords: Music performance; Still acquisition; Planning; Temporal control 1. Introduction Studies of expert and novice performance in many domains reveal both quantitative differences (such as accuracy and speed) and qualitative differences (such as * Corresponding author. Present address Laboratoire de Psychologie ExpeÂrimentale, UMR 8581 Universite Rene Descartes, Centre Universitaire de Boulogne, 71 avenue Edouard Vaillant, 92774 Boulogne-Billancourt, Cedex, France. E-mail address: drake@idf.ext.jussieu.fr (C. Drake) different error types that arise, and sensitivity to different structural aspects). Skilled performance in tasks such as typing, chess playing, and sports is usually faster, more accurate, and re¯ects more highly structured representations of the task than performance by less accomplished individuals (Ericsson & Smith, 1991; Ericsson & Staszewski, 1989). To a certain extent, these differences between experts and novices can be generalized to a complex task such as music performance. For example, skilled pianists can play faster, make fewer errors, and perform more consistently (McKenzie, Nelson-Schultz & Wills, 1983), and plan ahead (anticipate) more than beginners (Palmer & Drake, 1997). However, the structure of musical sequences imposes additional constraints on performance, namely temporal constraints, which have been severely underestimated in previous descriptions of skilled performance. Temporal constraints in music performance include the need to produce events in a ¯uent or continuous manner, maintain production rate or tempo, and preserve certain durational relationships between events. In this paper we examine the cognitive bases of music performance with an emphasis on the role of temporal constraints, and their relation to how performers plan, or prepare sequences for production. We approached the study of music performance in terms of skill acquisition. One reason for this choice stems from the fact that the temporal structure of music tends to be completely mastered in expert performance, providing few instances of temporal de®cits or breakdowns. As this study demonstrates, one major component of novices' music performance involves the learning of temporal structure, and performances by beginner musicians provide naturalistic conditions in which to observe temporal breakdowns. We report a study of how musicians of various skill levels learn to play a novel piece of music, focussing in particular on their ability to control both `what' they play and `when' they play it. Music performance offers an excellent domain for study of the cognitive processes responsible for both the production of serial order (which event to produce next) and event timing (when), because the temporal precision seen in music performance and the conceptual structures of rhythm and meter indicate that timing is integral to performance goals (for a review see Palmer, 1997). The acquisition of performance skills can be studied within several paradigms. One approach is to compare performances of musicians at varying levels of expertise (Drake, Dowling & Palmer, 1991; Ericsson, Krampe & Tesch-Romer, 1993). A second approach is to study the progressive acquisition of skills during the learning of a novel piece (Gruson, 1988). We combined these approaches in this study by examining how novice and expert pianists of varying skill levels learn to play a novel (unfamiliar) piece of music. We traced the process of skill acquisition in two ways: ®rst, we tracked improvements in both quantitative and qualitative characteristics of performance by comparing the frequency and type of errors for ®ve groups of pianists who vary in age, training, and sight-reading ability. Second, we examined the time-course of learning by studying improvements observed while each musician learns to play a novel piece of music over many repetitions. The combination of these two approaches provides an indication of which changes are due to general skill level

(long-term learning) and which are due to increased familiarity through 2 C. Drake, C. Palmer / Cognition 74 (2000) 1±32 speci®c practice of a novel musical piece (short-term learning). Although music performance usually entails much longer and less structured forms of practice, this early (®rst) stage of practice allows examination of differences in early learning stages that commonly re¯ect increased variability in other domains (Ivry, 1996; Newell & Rosenbloom, 1981). 2. Four issues of skill acquisition in music performance 2.1. Traditional measures of skill: ¯uency speed and accuracy Previous research has emphasized improvements in speed and accuracy with practice, both across and within skill levels (Chi, Glaser & Farr, 1988; Gruson, 1988; McKenzie et al., 1983), in tasks such as reading inverted text (Kolers, 1975) and cigar manufacturing (Crossman, 1959). The observed learning curves are usually best ®t by a power function, which show a rapid improvement at the beginning of practice, with a gradual slowing in improvement as amount of practice increases (Anderson, 1982; Crossman, 1959; Logan, 1988; MacKay, 1982; Newell & Rosenbloom, 1981; Schneider, 1977). Many theoretical interpretations have been suggested to explain the power function, including increases with learning in the size of, and relationships among, memory units or chunks (see Rosenbaum (1991) for a review). Skilled performance is thus often depicted in terms of the combination of speed and accuracy, or ¯uency, that results in a smoothness of production in which the correct actions are performed rapidly and without temporal interruption (MacKay, 1982; Shaffer, 1980). However, studies that use ¯uency as an indicator of skill have focussed primarily on tasks that require rapid, accurate performance, such as typing, tapping, and problem solving. The only temporal goal of these tasks is to perform as fast as possible while making as few errors as possible. A speed/accuracy trade-off is usually observed under these conditions: faster performances tend to contain more errors (MacKay, 1971; Newell & Rosenbloom, 1981). We examined whether speed and accuracy increase in music performance in a manner similar to the power law function. We traced improvements in terms of performance tempo (a measure of speed) and error rates (a measure of accuracy) at different skill levels (re¯ecting long-term practice) over multiple repetitions of the same piece (re¯ecting short-term practice). We also tested for a speed/accuracy trade-off. 2.2. Additional characteristic of skill: ¯uency relative timing and pitch accuracy Complex tasks such as music performance and speech impose additional task constraints beyond ¯uency, namely, temporal constraints. Overall speed or performance tempo is less important than producing events at the correct point in time. Musicians can create small expressive timing variations, but their productions must re¯ect the event duration categories speci®ed in a musical score in Western tonal music, which are usually simple integer ratio relationships. Performance is considC. Drake, C. Palmer / Cognition 74 (2000) 1±32 3 ered de®cient if it does not respect these duration categories. Therefore, it is not surprising that a speed/accuracy trade-off has not been observed in expert adult pianists' performances of memorized sequences under speeded conditions (Palmer & van de Sande, 1993) or under normal (unspeeded) conditions (Palmer & van de Sande, 1995): error rates did not increase in faster performances. The concept of ¯uency needs to be modi®ed in the case of music performance and other complex serial tasks with high temporal constraints, such as speech. We enlarge the concept of ¯uency in music performance to include the relative timing of events. A skilled music performance is ¯uent, in the sense that correct actions are performed at the correct moment in time relative to surrounding events. An additional characteristic of expertise in music performance therefore may be the ability to master temporal constraints by performing events at the correct time. Because speed is replaced by relative timing as a task goal in music performance, we predict a relative-timing/pitch accuracy trade-off, with performers able to produce either the relative timing at the expense of pitch accuracy or vice versa, instead of a speedaccuracy trade-off. 2.3. Mastering the temporal structure of music: temporal continuity, underlying beat and metrical structure

Western tonal music has a hierarchical temporal structure, in which event durations combine to form basic rhythmic patterns which in turn combine to form higherorder rhythmic patterns (Cooper & Meyer, 1960; Lerdahl & Jackendoff, 1983; Yeston, 1976). Three speci®c temporal constraints on how music is performed are examined here: temporal continuity, underlying beat and hierarchical metrical structure. The ®rst constraint, temporal continuity, re¯ects a requirement to produce events in a continuous fashion without hesitations or temporal interruptions. Skilled performers strive to respect the continuous nature of music, for example, concert pianists can adjust the performance of rapidly executed events to compensate for an error in the timing of earlier events (Shaffer, 1980). However, performances by novice musicians demonstrate frequent interruptions, despite music teachers' explicit instructions to continue playing. For example, novices frequently stop to correct errors, whereas advanced musicians tend not to correct their errors (Palmer & Drake, 1997). The second temporal constraint, maintaining an underlying beat, re¯ects the need to produce events at time intervals that maintain recognizable (but not necessarily exact) integer ratio relationships. Western tonal music is composed of event durations that are related by integer ratios: event durations are usually one, two, three, or four times as long as the shortest event duration (occasionally more). Respecting these integer ratios is thought to be accomplished by the perceptual extraction of an underlying beat (Dowling & Harwood, 1986; Drake, 1998; Parncutt, 1993). Sensitivity to durational relationships functions early in life: children over 5 years old are better able to reproduce rhythms constructed around simple integer ratios (1:1 and 1:2) than rhythms with more complex integer ratio relationships (Drake & GeÂrard, 4 C. Drake, C. Palmer / Cognition 74 (2000) 1±32 1989). One limitation in generalizing to naturalistic music performance from these simple tasks, however, is the simplicity of the sequences, which usually contained only ®ve to seven events. The third temporal constraint, metrical structure, re¯ects the fact that Western tonal music contains temporal regularities de®ned over longer time-spans (spanning many beats). Meter refers to a pattern of periodic accent placement that arises from alternating strong and weak beats. People reproduce event sequences that are consistent with a metrical framework (those that are constructed around integer ratios) more correctly than those that violate a metrical framework (Essens, 1986, 1995; Essens & Povel, 1985; Povel, 1981), and the same pattern of event durations (such as short-short-long) is performed with different relative timing when placed in different metrical contexts (Clarke, 1985). Music performances often contain expressive nuances that re¯ect the metrical structure (Drake & Palmer, 1993; Sloboda, 1983; Shaffer, Clarke & Todd, 1985). Events aligned with metrical strong beats are often lengthened (Gruson, 1988) and played louder (Sloboda, 1983). Errors in child pianist's well-learned performances indicate sensitivity to meter: more corrections occur between metrical measures than at other metrical positions (Palmer & Drake, 1997). Some research suggests that young children are not as sensitive to metrical structure as adults (Bamberger, 1980; Drake, 1993a,b; 1997; Drake & GeÂrard, 1989; Smith, 1983); however, few have compared childrens' and adults' sensitivity in the domain of music performance. We addressed the effects of both general skill and speci®c practice on performers' ability to respect these temporal constraints of music. Maintaining continuous temporal performance, respecting the underlying beat, and sensitivity to metrical structure may improve with both general skill and speci®c practice. 2.4. Preparation of movement sequences: the ability to plan and the range of planning Many theories of sequence production assume that performers construct a mental plan prior to performance that speci®es the content and ordering of sequence events (Keele & Summers, 1976). The number of events that can be prepared simultaneously is limited by memory constraints, and long sequences are thought to be partitioned into shorter subsequences (or segments) during the construction of mental plans (van Galen & Wing, 1984. Evidence from timing and errors in music performance also suggests that musical

sequences are partitioned during planning into shorter segments (Palmer & van de Sande, 1995). Planning abilities are often re¯ected in anticipatory or future-oriented behavior. For example, production errors in speech and music performance often reveal events that are anticipated (produced ahead of time). A comparison of beginner and intermediate children's piano performances indicated increased anticipatory behavior: children with 2±3 years more musical training showed more anticipatory relative to perseveratory errors (Palmer & Drake, 1997). Speci®c practice also leads to more anticipatory behavior. Speakers producing tongue-twisters showed an increased proportion of anticipatory errors with practice, and the anticipatory proportion of C. Drake, C. Palmer / Cognition 74 (2000) 1±32 5 errors was predicted by the overall error rate (Dell, Burger & Svec, 1997). Dell et al. (1997) propose a model in which practice affects the strength of connections in mental plans to both present and future events but not to past events. Thus, planning capacities improve with practice as a function of increased memory associations among sequence elements. Whether these predictions of speci®c practice apply to other sequence production domains such as music performance is unknown. A related question that arises is the range or scope of planning during sequence production. Many errors, such as substitutions, involve the production of a sequence element in a different serial position from that intended by the performer. These errors indicate an interaction between non-adjacent elements that must be simultaneously accessible during planning of the sequence (Garrett, 1980). We assume here that the range of planning, or distance in number of sequence elements between the interacting items (an error and its assumed source), indicates which elements are simultaneously accessible at any given time. Constraints on the range of planning are evidenced in speech errors, whose interacting elements tend to span structural units (Garrett, 1980). Evidence from music performance errors also indicates structural constraints on the range of planning; phrase structure and serial distance limit the range over which sequence elements interact (Palmer & van de Sande, 1993, 1995), similar to error patterns found in speech (Garcia-Albea et al., 1989). Furthermore, the range of planning, as re¯ected in production errors, increases from beginner to intermediate levels of music performance (Palmer & Drake, 1997). Whether general skill level or speci®c practice affect the range of planning in terms of greater serial distances, larger structural units, or other factors, is unknown. We address the effects of both general skill level and speci®c practice on planning capacities underlying music performance. Both general skill level and speci®c practice should lead to mental plans that are more future-oriented (as measured in anticipatory errors), and incorporate events from larger ranges of the sequence (as measured by interacting sequence events in errors). 3. Experimental study of musical skill acquisition We traced the progress of ®ve groups of pianists varying in age, amount of musical training, and sight-reading ability as they learn a novel musical piece. Each pianist performed a piece of music appropriate to his/her skill level in eleven practice trials. We examined the performance tempo, as well as the number and type of errors (pitch, duration, corrections, and pauses), by comparing each performance with intended events notated in the musical score. Based on the previous literature, we make the following predictions. First, improvements in both error rates (an indication of accuracy) and performance tempo (an indication of speed) should be evidenced over skill levels and practice trials, probably following a power function. We do not however expect a speed/accuracy trade-off as is usually observed in motor tasks, because temporal constraints other than speed (i.e. relative timing) in¯uence the timing of music performance. Second, we do expect a relative-timing/pitch accuracy trade-off in 6 C. Drake, C. Palmer / Cognition 74 (2000) 1±32 music performance. Beginner pianists may only be able to perform correct pitches at the expense of timing, whereas more skilled pianists may be able to incorporate both pitch and timing dimensions. We compared the relative frequency of pitch and timing errors as a function of skill level and speci®c practice. Third, skilled performers may respect

the temporal structure (relative timing of events) more than beginners. The ability to maintain temporal continuity, respect the underlying beat, and sensitivity to the hierarchical metrical structure may also improve with general skill level and specific practice. Finally, we expected that both general skill and specific practice should lead to mental plans that are more future-oriented (as measured in anticipatory proportions of errors) and incorporate events from larger ranges of the sequence (as measured by interacting sequence events). 3.1. Method 3.1.1. Participants Five groups of 12 pianists (total of 60) were recruited through local music teachers' associations in the Columbus, Ohio music community: 48 pianists were children and 12 pianists were adults. Children were between 7 and 16 years of age, and adults were between 20 and 30 years of age. Pianists were assigned to one of five groups based on their age and sight-reading ability, as described in the Procedure (Section 3.14) below. 3.1.2. Stimulus materials Three short Western tonal musical pieces which varied in difficulty were composed for the experiment by a musician to ensure that they would be unfamiliar. The first four measures of each piece are shown in Fig. 1. All three pieces were composed in a Western classical style, with simple meter, melody, and accompaniment. Piece 1, the easiest piece, was a single-voiced melody of eight measures, in C. Drake, C. Palmer / Cognition 74 (2000) 1±32 7 Fig. 1. First four measures of the three musical pieces used in this experiment. two four-measure phrases. Pieces 2 and 3 were composed of a melody and two accompanying voices: piece 2 had a chordal accompaniment and was composed of two four-measure phrases, and piece 3 was based on a French folk-tune, composed in an ABA format, with each section lasting 16 measures. The pieces are referred to hereafter as p1, p2 and p3. 3.1.3. Apparatus Pianists performed the pieces on a computer-monitored upright Yamaha Disklavier acoustic upright piano. Optical sensors and solenoids in the piano allowed precise recording and playback without affecting the touch or sound of the acoustic instrument. The timing resolution was 2 ms for note events, with precision (measured by the standard deviation of onset-to-onset durations during recording) within 0.8% for durations in the range of the performances. Errors were identified by computer comparisons of recorded pitch and duration values with the information given in the musical scores. 3.1.4. Procedure and design The pianists (and their parents) were first interviewed about the pianists' musical background. They performed a familiar musical piece of their own choosing, to become acquainted with the piano. Then they performed one of the unfamiliar musical pieces. All musicians except the adults and the most advanced children began with the easiest piece (piece 1). If they were able to perform the piece with no errors after two trials, they were given the next most difficult piece to try. This procedure was repeated until the musicians settled on a piece of appropriate difficulty. Once the appropriate piece to be learned was established, the pianists performed it five times. Then they performed a short unrelated task, introduced to prevent boredom (an issue for the youngest pianists). Then the pianists were asked to perform the piece once more and that performance was played back by computer for them to hear. Instructions were then given to emphasize either: (1) the meter, (2) the phrase structure marked on the score, or (3) what the music made them think of. These instructions had no significant effects on results presented here and thus are dropped from further analyses. Each pianist then performed the piece an additional five times, with these instructions. In all, each pianist performed one of the pieces 11 times. At the end of the experiment, the children chose a small gift from a collection prominently displayed in the laboratory, and the adults (or parents) received $8. 3.1.5. Skill classification Each pianist was first classified into one of three sight-reading levels depending on piece difficulty (p1 easiest, p2 intermediate, p3 hardest). Child pianists were further classified on the basis of their age (`young' 7±11 years old, `old' 12±16 years old). This assignment resulted in five groups of 12 pianists each, shown in Table 1: p1-young, p2-young, p2-old, p3-old, and p3-adults. A design that crossed age and sight-reading (skill) level completely was not possible

because the easiest piece was too easy for the older pianists to perform with any errors, and the hardest piece was too dif®cult for the younger pianists (see Table 1). This design does 8 C. Drake, C. Palmer / Cognition 74 (2000) 1±32 however allow the study of improvements during learning due to age, training, and sight-reading ability. Mean age and years of training (private piano instruction) for each group are shown in Table 2. The `old' groups (p2-old, p3-old) were signi®cantly older than the `young' groups (p1-young, p2-young: F 1; 55 74:6, P , 0:01) and the adults were signi®cantly older than the old groups: (F 1; 55 487:3, P , 0:01). The two `old' groups had signi®cantly more years of training than the two `young' groups (F 1; 55 31:5, P , 0:01), and fewer years of training than the adults (F 1; 55 109:3, P , 0:01). There were no signi®cant differences in mean age or training between the two `young' groups (p1-young and p2-young) or between the two `old' groups (p2-old and p3-old). 3.1.6. Error classi®cation Errors were identi®ed by computer and coded according to a system similar to those used with speech errors (Dell, 1986; Garrett, 1975), adapted previously for the music performance domain (Palmer & Drake, 1997; Palmer & van de Sande, 1993, 1995). Five types of errors were identi®ed in the performances: pitch errors, duration errors, pitch/duration combinations, corrections, and pauses. Examples of the error types are shown in Fig. 2. The experimenters veri®ed the computer coding of the errors based on two musician listeners' notated transcriptions of all of the performances. The musicians' coding was adopted in the rare cases of disagreement with the computer output (usually concerning the position of corrections), which occurred most often in the single-voice piece (p1-young performances) for which C. Drake, C. Palmer / Cognition 74 (2000) 1±32 9 Table 2 Mean, standard deviations, and range of age and training (in years) for the ®ve groups Group Age Training Mean SD Range Mean SD Range P1-young 9.1 1.2 7±11 3.5 1.5 1.5±5.5 P2-young 10.3 1.1 8±11 4.7 1 1.5±7 P2-old 13.1 1.1 12±15 6.6 2.1 4±10 P3-old 14 1.2 12±16 8.4 1.6 5±11 P3-adults 25.7 2.6 20±30 15.3 3.1 10±20 Table 1 The ®ve groups included in this study Age Dif®culty of musical piece Easiest (piece 1) Intermediate (piece 2) Hardest (piece 3) Young (7±11 years) p1-young p2-young Old (12±16 years) p2-old p3-old Adults (.18 years) p3-adults error position codings could be ambiguous due to the lack of accompanying voices. Because pause and duration errors are de®ned here for the ®rst time, the error de®nitions are described in detail below. Pitch errors were de®ned as events whose pitch component differed from the notated pitch information. Pitch errors were coded on the same dimensions as previously (see Palmer & van de Sande, 1993, 1995 for details): type (additions, deletions, substitutions, or shifts), source (contextual, non-contextual, or combination), size (chord, note, or note-chord combination), and movement direction (anticipatory, perseveratory, or exchange). Contextual errors are those whose pitch matches the notated (intended) pitch of a position adjacent to the error. In addition, some errors have a directional component; they can be anticipatory (pitches intended for future positions) or perseveratory (intended for past positions). All errors were coded in the most parsimonious way (to minimize the total number of distinct errors). Following this strategy, multiple pitch errors whose onsets occurred within a small temporal window (75 ms for adults and 100 ms for children) were coded as chord error size; single-note errors were coded as note error size. Duration errors were de®ned as events whose performed interonset duration differed from the duration notated in the musical score. (In this paper, duration 10 C. Drake, C. Palmer / Cognition 74 (2000) 1±32 Fig. 2. Examples of error coding system. refers to the interonset interval de®ned by two successive events.) The observed duration of each event was compared with its predicted duration (based on the duration of the nearest preceding correct event), in order to incorporate local tempo changes. An event duration was considered in error if the observed duration differed from the predicted duration by more than 50% (thereby entering the next duration category in musical nomenclature). The observed duration of the

error was then classi®ed into the nearest notational duration category (eighth-notes, dotted eighth-notes, quarter notes, etc.) based on the ratio of observed/predicted durations. In addition, each duration error was coded for beat disruption: duration errors were considered beat-disruptive if the ratio of the observed (error) to the predicted event duration was not an integer multiple of the beat duration (the beat duration equaled one quarter-note in each musical piece, as de®ned by the denominator of the time signature). Thus, a quarter-note event performed in error as a dotted quarternote would be coded as beat-disruptive, whereas a quarter-note event performed in error as a half-note would be coded as non-beat-disruptive. In the case of two adjacent duration errors that occurred within the same beat period (quarter-note), the beat disruption for both errors was coded on the basis of their summed observed duration relative to the (summed) predicted duration. Pitch-duration errors were de®ned as events for which both the pitch and duration components were in error. The pitch and duration components of pitch/duration errors were each coded in the same manner as individual pitch errors and duration errors. Pauses were de®ned as interonset intervals greater than three times the predicted duration, resulting in values often greater than one second (similar to classi®cations of hesitation pauses in speech; Fodor, Bever & Garrett, 1974). They may represent silence or a held tone. Predicted durations were estimated from the nearest preceding (correct) event duration, as for duration errors. Pauses were coded as occurring between the serial locations of the event onsets preceding and following the pause. Corrections were de®ned as a series of (error) additions of more than one event, followed by a pause, and then a (correct) restart, following previous studies of music performance errors (Palmer & Drake, 1997). Corrections were coded as occurring between the event locations preceding and following the correction. Learning errors. Some errors in sight-reading tasks may arise from perceptual (reading) or input causes, as well as from production sources (cf. Lehmann & Ericsson, 1993). To separate these sources, we de®ned learning errors as those pitch or duration errors that occurred consistently in the same event location in four or more of the ®rst ®ve performances. Pitch learning errors formed less than 10% of the total pitch errors and were excluded from the following analyses. In contrast, duration learning errors formed 63% of the total duration errors, 99% of which occurred in performances of piece 1 by the least skilled group (p1-young). Subsequent analyses of these duration errors indicated that some of the least skilled pianists in the P1-young group misinterpreted the time signature as 6/8 instead of the notated 4/4 signature, and performed the event durations to match the incorrect time signature. Thus, the intention rather than the execution was incorrect in these errors. All duration and pitch learning errors were excluded from further analyses. C. Drake, C. Palmer / Cognition 74 (2000) 1±32 11 Error rates. Statistical analyses were conducted on error rates (computed as number of errors relative to number of error opportunities) for each pianist's performances, to adjust for the different lengths of the musical pieces. There was a total of 6089 errors in the 60 (pianist) £ 11 (trial) performances (p1-young 1685, p2-young 1277, p2-old 983, p3-old 1669, p3-adult 475). Error rates were computed from a combination of pitch, duration, correction, and pause errors, relative to chance estimates for each error type. Pitch error rates were de®ned as: [# single note pitch errors 4 number of single pitches in piece, summing across chords and single-notes] 1 [# chord errors 4 number of events in piece]1 . Duration error rates were de®ned as: # duration errors 4 (number of events in piece 2 1), because there is no event onset to determine the (interonset) duration of the last event. Pitch/ duration errors (those containing both a pitch and duration component in error) were divided equally (assigned 0.5) among pitch and duration error rates. Pause error rates were de®ned as: [# pauses 4 total events in piece 2 1), and correction error rates were de®ned as: [# corrections 4 total events in piece 2 1]. The mean error rate for each performance was computed from the average of these four numbers. Analyses across all error types were conducted on error rates, and analyses that compare speci®c error types were

conducted on percentage of total errors. 4. Results 4.1. Traditional measures of performance: accuracy and speed We ®rst examined two hallmarks of skilled performance: increased accuracy and speed, as measured by error rates and performance tempo. 4.1.1. Accuracy As shown in Fig. 3a, error rates decreased over the ®ve groups and eleven practice trials. An ANOVA on error rates by group (®ve), practice trials (11), and error type (four) revealed a signi®cant effect of group ($F_{4, 55} = 20.7$, $P < 0.01$), with lower error rates for the more skilled groups, and of practice trial, ($F_{10, 550} = 19.9$, $P < 0.01$), with error rates decreasing over practice. A signi®cant interaction between practice trial and group ($F_{40, 550} = 2.7$, $P < 0.01$) also indicated that the rate at which a new piece improved in accuracy across the eleven trials varied 12 C. Drake, C. Palmer / Cognition 74 (2000) 1±32 1 Pitch error rates can be computed for single-note errors and chord errors (those involving multiple simultaneous note events) by several methods: (1) combine chord and single-note errors and compare total with chance estimate of number of event locations in sequence; (2) treat chords as multiple single-note errors, combine them with number of single-note errors and compare total with chance estimate of total number single pitches in sequence; or (3) compare number of single-note errors with chance estimate of total number of single pitches and number of chord errors with chance estimate of number of event locations; then add these two ratios (each error is represented only once in the two ratios). The ®rst method is biased toward lower error rates when chord errors are uncommon, the second method is biased toward higher error rates when chord errors are uncommon, and the third method typically results in an intermediate value. Given no a priori reason to expect differences in chord versus single-note errors, we use the third method, which represents a balance of the two potential biases. across the groups. To analyze this interaction further, separate ANOVAs were conducted for each group on trials 1±5 and trials 7±11. These comparisons indicated signi®cant improvement over both the ®rst ®ve and last ®ve trials for p1-young (Dunn±Bonferroni, $P < 0.05$), only over the ®rst ®ve trials for groups p2-young, p2-old, and p3-old ($P < 0.05$), and borderline signi®cance over the ®rst ®ve trials for p3-adults ($P = 0.055$). Greatest improvements during practice were found in the least skilled groups. 4.1.2. Tempo We also examined changes in tempo with learning. The initial tempo at which the pianists performed the pieces was estimated by the mean quarter-note beat duration (IOI) of events in the ®rst error-free measure for each performance (the quarter-note represents one beat, as indicated by the notated time signature[2]). This measure of initial tempo is shown in Fig. 3b over practice trials for each group. An ANOVA on the mean quarter-note IOIs for the ®ve groups and 11 trials indicated that the mean C. Drake, C. Palmer / Cognition 74 (2000) 1±32 13 Fig. 3. Improvements over eleven practice trials for the ®ve groups for mean error rates (a) and mean initial IOI (b). 2 We report initial tempo measures rather than average tempo measures because correction errors and pauses, which occurred at different rates over groups, distorted sequence duration measures. tempo increased (performers played faster) over the ®ve groups ($F_{4, 55} = 20.2$, $P < 0.01$). There was also a signi®cant increase in tempo over the eleven practice trials ($F_{10, 550} = 18.6$, $P < 0.01$) and a signi®cant interaction between group and practice trial ($F_{40, 550} = 2.0$, $P < 0.01$). To address this interaction, separate ANOVAs on trials 1 through 5 and trials 7 through 11 were conducted for each group. Performance by all groups except p1young became signi®cantly faster over the ®rst ®ve trials (Dunn±Bonferroni, $P < 0.05$). Only p3-old and p3-adults showed increased tempo over the last ®ve takes ($P < 0.05$). We tested whether changes in accuracy and speed over practice followed a power function. Regression analyses predicting both error rates and performance tempo from practice trial

were conducted for each performer on both log (base 10) and linear values (log/log and lin/lin), as shown in Table 3. Both functions ®t the error rates and tempo values well; however, a power function accounted for more variance than a linear function for all groups for the error rates and for all groups except p1young for the performance tempo. Thus, performance tempo and accuracy improved both with skill level and practice. 4.1.3. Speed/accuracy trade-off To check whether the changes error rates and initial performance tempo shown in Fig. 3 re¯ected a speed/accuracy trade-off, the two were correlated across performances: a negative correlation would indicate more errors at faster tempi. In contrast, positive correlations were found between mean initial IOI measures and error rates within performances, both within practice trials across skill levels (ranging from r 0:62 to 0.76, P , 0:01) and across practice trials within skill levels (ranging from r 0:15 to 0.55, P , 0:10). Although there was limited change in performance tempo overall, the slower performances tended to contain more errors, contrary to predictions of a speed-accuracy trade-off. 4.2. Relative-timing/pitch accuracy trade-off Some error types re¯ect failures in accuracy (such as pitch errors), whereas others re¯ect failures in relative timing (such as duration errors). In this section we examine 14 C. Drake, C. Palmer / Cognition 74 (2000) 1±32 Table 3 Regression analyses (r): mean of individual pianists within group (N 11) Error rates Tempo Mean lin/lin Mean log/log Mean lin/lin Mean log/log p1-young 2 0.72a 2 0.79b 2 0.76b 2 0.71a p2-young 2 0.89b 2 0.94b 2 0.74b 2 0.92b p2-old 2 0.92b 2 0.92b 2 0.90b 2 0.96b p3-old 2 0.84b 2 0.95b 2 0.92b 2 0.98b p3-adult 2 0.53 2 0.59 2 0.83b 2 0.99b a P , 0:05: b P , 0:01: improvement in four error types (pitch, duration, pause, and correction errors) over skill levels and practice trials. 4.2.1. Improvement in each error type over groups and practice trials The main ANOVA on error rates revealed, in addition to the previous main effects of group and practice trial, a main effect of error type (F

3; 165 31:5, P , 0:001): pauses were the most frequent type of error (50.1% of all errors across groups), pitch and correction errors were intermediate (pitch 19.5%, corrections 17.5%) and duration errors were least frequent (12.6%). More important, there was a C. Drake, C. Palmer / Cognition 74 (2000) 1±32 15 Fig. 4. Mean error rates by error type and group. two-way interaction between group and error type (F

12; 165 7:9, P , 0:01). As shown in Fig. 4, the least skilled groups (p1-young, p2-young, p2-old) made more pauses, corrections, and duration errors, whereas the most skilled groups (p3-old, p3-adults) made mainly pitch errors. There was also a signi®cant two-way interaction between practice trials and error types (F

30; 1650 5:5, P , 0:01) and a three-way interaction between group, practice trials, and error types (F

120; 1650 1:6, P , 0:01). Although error classi®cations may limit generalizations about distinctions among error types, the general pattern was that all groups except the adults showed improvements over practice in pauses and corrections (the predominant error types), whereas the adults showed improvements only in pitch error rates (the only predominant error type, as shown in Fig. 4). 4.2.2. Relative timing versus pitch accuracy We examined the relationship between relative timing and pitch accuracy across skill levels. The percentage of relative timing errors (duration and pause errors) and pitch errors relative to total errors (pitch, duration, pauses and corrections) was computed for each performance, and is shown for each group in Fig. 5a. An ANOVA on these percentages by group, error type (pitch or time), and practice trial (1±11) indicated a signi®cant interaction of group with error type (F

4; 55 18:9, P , 0:01). The less skilled groups (p1-young, p2-young, p2- old) made a higher proportion of timing errors than pitch errors, suggesting that they were concentrating primarily on playing the correct notes irrespective of temporal constraints. The more advanced pianists (p3-old, p3-adults) showed the opposite pattern, with a higher proportion of pitch errors to timing errors. Thus, beginners concentrated primarily on the `what' to the

detriment of the `when', and advanced pianists mastered the `when' and only erred on `what' should be performed. There was also a signi®cant interaction of error type with practice trial (F

10; 550 2:80, P , 0:01). As shown in Fig. 5b, the proportion of pitch errors increased and timing errors decreased over practice for all groups. Thus, the tendency to err in time shifted to a tendency to err in pitch within the eleven-trial practice session, and across skill levels. An overall negative correlation between pitch and timing error percentages on individual performers' data con®rmed this relative-timing/pitch accuracy trade-off (r 20:81, P , 0:01). 4.2.3. Tests of independence between pitch and time components A tradeoff between what pitch events occur and when across the musical sequence suggests that pitch and timing components of individual units also may not be produced independently. To assess their independence, we compute the probability of joint pitch/duration (P/D) errors based on the error rates for the separately occurring pitch (P) errors and duration (D) errors [prob(P) £ prob(D)]. As shown in Table 4, P/D errors were observed more often than expected from the individual estimates for all skill groups. For example, the p1young group had twice as many joint P/D errors as expected. Interactions were largest for the most skilled groups; however, even the least skilled group (who produced the most duration errors) showed signif16 C. Drake, C. Palmer / Cognition 74 (2000) 1±32 C. Drake, C. Palmer / Cognition 74 (2000) 1±32 17 Table 4 Observed and predicted pitch/duration error rates (number per hundred events) and their ratio for the ®ve groups Pitch/duration errors (observed) Pitch/duration errors (predicted) [prob(P) £ prob(D)] Observed/predicted ratio p1-young 1.48 0.700 2.1 p2-young 0.97 0.033 29.4 p2-old 0.97 0.026 37.3 p3-old 0.19 0.003 63.3 p3-adults 0.04 0.0002 200 Fig. 5. Percentage pitch and timing errors for the ®ve groups (a) and 11 practice trials (b). icantly more interaction errors than expected. These ®ndings suggest that the pitch and duration components interact even at the individual unit level of performance. Pitch/duration errors could arise from an attentional failure that affects both dimensions when concentration is lost, or from a memory failure that re¯ects the retrieval of another unit whose pitch and duration components both differ from those intended. To evaluate these possibilities, we compared whether the pitch and duration components of errors tended to arise from the same source; such an outcome would be consistent with memory retrieval of a different unit whose pitch and duration components were encoded together. Because pitch/duration errors that contained a possible intruder unit (additions and substitutions) formed a small subset of the total errors (n 118), this analysis was conducted over groups. 64.4% of these errors had pitch and duration components that came from the same identi®able source (an event intended for nearby in the sequence), signi®cantly more often than a conservative chance estimate of 50% (binomial test, P , 0:013). These ®ndings provide converging evidence that pitch and duration components interact in sequence production, and are probably encoded and retrieved from memory together. 4.3. Mastering temporal structure: temporal continuity, underlying beat and metrical structure In this section, we further examine errors in music performance as indicators of sensitivity to speci®c temporal dimensions of music (continuity, beat, and meter). 4.3.1. Temporal continuity Performers' ability to maintain temporal continuity was measured in terms of the percentage of total errors that disrupted the temporal ¯ow (pauses, corrections, and beat-disruptive duration errors). Three performers (in p3-adults) were excluded from the analysis because they had no pauses, corrections, or beat-disruptive duration errors. An ANOVA performed on the percentage of temporally-disruptive errors by group and practice trial indicated a main effect of group (F

4; 52 20:58, P , 0:01); the ability to respect the temporal ¯ow increased with skill level. There was also a main effect of practice (F

10; 520 3:60, P , 0:01): the ability to respect the temporal ¯ow improved with practice for all groups. Fig. 6 shows the percentages of errors disrupting continuity across groups and

practice trials, averaged for the ®rst half of the trials (1±5) and the last half (7±11). There was no interaction of group and practice trial. Thus, performers demonstrated increased ability to maintain temporal continuity with both skill level and speci®c practice. 4.3.2. Underlying beat A more speci®c dimension of temporal structure in music performance is main18 C. Drake, C. Palmer / Cognition 74 (2000) 1±32 3 A chance estimate of 50% implies that pitch and duration components have only two possible error outcomes; the musical pieces contained many more pitch and duration alternatives than two, and thus 50% is a conservative chance estimate that nevertheless yields signi®cant results. tenance of the underlying beat. Some timing errors do not disrupt the musical ¯ow but do disrupt the underlying beat. For example, a quarter-note replaced by a halfnote does not disrupt the underlying beat (the quarter-note), but a quarter-note replaced by a single eighth-note does disrupt the underlying beat. The percentage of beat-disruptive errors relative to total number of duration errors was computed for each pianist across trials; there were too few errors to test differences across trials. The percentage of beat-disruptive errors decreased with skill; p1-young 40.4%, p2-young 50.3%, p2-old 35.7%, p3-old 24.7%, p3-adults 31.3%. Group data were combined within piece (p2-young and p2-old; p3-old and p3-adults) to provide enough data for group comparisons; beat-disruptive errors were signi®- cantly lower than the chance estimate of 50% (probability of the error resulting in an integer multiple of the beat) only for the highest skill group (p3 26.5%; t(14) 2 2.71, P , 0:05). Thus, the ability to respect the underlying beat of the music may improve with skill level. 4.3.3. Metrical structure We tested performers' sensitivity to meter by examining the position of each error type relative to the metrical boundaries indicated by vertical barlines separating measures. Previous evidence from music performance indicated that corrections were more likely to occur at metrical boundaries than at other positions (Palmer & Drake, 1997). ANOVAs on the mean error rates by metrical position and error type indicated a signi®cant interaction between metrical position and error type for each group (P , 0:05). Only correction errors indicated effects of metrical position. Therefore, we compared the percentage of total corrections that occurred at metrical barlines for each group, relative to a chance estimate of how often errors were expected to occur at barlines in each piece (number of notated events at metrical C. Drake, C. Palmer / Cognition 74 (2000) 1±32 19 Fig. 6. Percentage ¯ow-disruptive errors for trials 1±5 and 7±11 for the ®ve groups. barlines, relative to total notated events). As shown in Fig. 7, corrections occurred signi®cantly more often at barlines than at any other metrical position for all groups (P , 0:01) except p1-young, the least skilled group. Thus, sensitivity to metrical structure was observed early and improved with skill level. One possible explanation for the increased frequency of corrections at barlines is the physical layout of the musical notation. Performance may be interrupted more often at locations that require large eye movements, such as line endings (which coincide with some metrical barlines). Error percentages and chance estimates for corrections were recomputed, excluding line breaks and events in the metrical bar preceding them. Again, corrections occurred more often at barlines than at other positions for all groups except p1young, relative to chance estimates (P , 0:01), similar to the results shown in Fig. 7. Therefore, it was unlikely that the metrical effects were attributable to eye movements alone. 4.4. Preparation of movement: anticipatory behavior and range of planning Here we examine performers' ability to prepare the contents (pitch) of events for production in terms of future-oriented (anticipatory) behavior and the scope or range of planning. 4.4.1. Anticipatory behavior We ®rst test differences among skill levels in anticipatory behavior by computing the proportion of anticipatory errors relative to anticipation and perseveration errors combined, for all contextual pitch errors (those in which the intruding event may have been intended for an adjacent position in the sequence). Only errors with an identi®able source in the immediately surrounding context were included; Piece 1 (a 20 C. Drake, C. Palmer / Cognition 74 (2000) 1±32 Fig. 7. Predicted and observed

percentage corrections at metrical barlines for the ®ve groups. single-voiced melody) was not included due to dif®culties in coding error movement for single-voice pieces with repeating pitches. An ANOVA on the proportion of anticipatory errors relative to anticipatory plus perseveratory errors (AP) by group (4) indicated a signi®cant effect (F 3; 44 3:4, P , 0:05): anticipatory proportion for p2-young 50.9%; p2-old 42.1%; p3old 62.8%; p3-adults 62.3%. Thus, the proportion of anticipatory errors increased with skill level, indicating increased future-oriented planning. We also examined changes in anticipatory proportions across practice. Because these movement errors formed a subset of the overall error data, errors were pooled across groups within trials to examine practice effects. According to Dell et al. (1997) model of sentence production, anticipatory proportion should increase because practice enhances the activation of the present and future relative to the past, and perseverations become less common relative to anticipations as performance improves with practice. The anticipatory proportion increased with practice from the ®rst two trials (AP 0:55) to the last two trials (AP 0:68). A regression analysis, predicting AP values by practice trials, indicated that AP increased significantly with practice over the last ®ve trials (r 0:91, P , 0:05) and a positive but not signi®cant trend in increased anticipatory behavior over the ®rst ®ve trials (r 0:57, P . 0:05). Thus, anticipatory behavior increased with practice. We next examined factors affecting the relationship between anticipatory proportion and error rate. The Dell et al. (1997) model predicts an inverse relationship between anticipatory proportion and overall error rate (log units), arising from a combination of practice, production rate, activation rate, and decay parameters. Following Dell et al. (1997), we computed the correlation between overall AP and overall pitch error rate (log units) for errors combined across the two Piece 2 C. Drake, C. Palmer / Cognition 74 (2000) 1±32 21 Fig. 8. Regression of total anticipatory proportion on log total pitch error rates for piece 2 and piece 3, over the 11 practice trials. groups (young and old) and two Piece 3 groups (old and adult), to combine the relatively small error subsets. As shown in Fig. 8, AP increases linearly as log total error rates increase for both groups and practice: r 20:64, P , 0:01. Each practice trial within group results in an AP along the regression line, with more advanced performers (p3) and trials representing more practice (11) falling toward the futureoriented (left/high) end of the line. This ®nding extends the general anticipatory effect (Dell et al., 1997); performances representing higher skill levels and more practice displayed fewer pitch errors overall and more anticipatory behavior. 4.4.2. Range of planning Another measure of planning ability is the scope or range of planning, re¯ected in the distance between interacting elements in pitch errors (Palmer & Drake, 1997). The standard deviation of error distances (in number of events) was computed for each group. Only errors with an identi®able source in the surrounding context, de®ned by the length of the shortest piece (30 events), were included. Bartlett's test of homogeneity of variance revealed signi®cant increases in the standard deviations across the ®ve groups (p1-young 2.43 events; p2-young 2.66; p2-old 3.36; p3-old 7.04; p3-adults 7.30; chi-squared (4) 642.8, P , 0:01). Because events can re¯ect different durations across pieces, error distances across groups were also compared in quarter-note units (the basic beat unit in all of the pieces)4 . Again, the error distances increased with skill level (p1-young 2.20 quarter-notes; p2-young 2.66; p2-old 3.36; p3-old 5.91; p3-adults 6.13; chi-squared (4) 508.1, P , 0:01). In addition, there were signi®cant differences between p2-young and p2-old (chi-squared (1) 13.3, P , 0:01); the older group demonstrated a greater range of planning than the younger group performing the same music. Interacting elements in errors tended to span larger distances as skill level increased, indicating that range of planning increases with skill. Piece-speci®c factors, such as pitch repetition rates, may in¯uence the identi®cation of error sources and their distances. Therefore, we computed a chance estimate for each piece based on the mean distance (in number of quarter-notes) between repeating pitches in the musical notation. Analyses

conducted only on errors whose distances fell outside the chance estimates con®rmed signi®cant differences in standard deviations across groups (p1-young 4.02; p2-young 8.62; p2-old 9.65; p3-old 9.11; p3-adults 9.16; chi-squared (4) 38.6, P , 0:01). Thus, the increased range of planning across skill levels was not due solely to differences in piece characteristics. Finally, we tested whether disruptions in timing of music performance coincide with disruptions in planning capacities. Temporal disruption errors (mean proportions of duration errors and pauses) and range of planning measures (standard deviations of pitch error distances) re¯ect independent dimensions of performance, as measured by different error characteristics (rates of timing errors and source of 22 C. Drake, C. Palmer / Cognition 74 (2000) 1±32 4 Error distances computed in ms-values across groups did not show an increase with skill level, perhaps due to global tempo differences and rates of temporal disruption across groups, both of which affect durational estimates of error distances. pitch errors). The two measures were correlated across groups. As shown in Fig. 9, increased skill levels showed both lower proportions of temporal disruptions and a greater range of planning (r 20:98, P , 0:01). As performers' access to sequence events spanned larger ranges, the timing of events was less likely to be disrupted. 5. Summary of ®ndings 5.1. Traditional measures of skill: speed and accuracy 2 Both performance tempo and accuracy improved with skill level and practice in early learning trials, consistent with predictions of a power function . 5.2. Additional characteristic of skill: ¯uency relative timing and pitch accuracy 2 No speed/accuracy trade-off was observed at any skill level or point in practice. 2 A relative-timing/pitch accuracy trade-off was observed, with a shift from focussing on pitch toward mastering time across both skill levels and practice. 2 Joint pitch/ duration errors occurred more frequently than expected from individual components. C. Drake, C. Palmer / Cognition 74 (2000) 1±32 23 Fig. 9. Correlation between mean timing error percentages and mean range of planning for the ®ve groups. 5.3. Mastering the temporal structure of music: temporal continuity, underlying beat and metrical structure 2 The ability to maintain temporal continuity improved with skill level and practice. 2 The respect of the underlying beat improved with skill level. 2 Sensitivity to metrical structure was observed early and improved with skill level. 5.4. Preparation of movement sequences: the ability to plan and the range of planning 2 The proportion of anticipatory errors increased with skill level and practice. 2 Range of planning (error distances) increased with skill level. 2 Range of planning increased as temporal disruptions decreased, with increased skill level. 6. Discussion This study is the ®rst to rigorously track cognitive changes that result from longterm and short-term learning in music performance. We have documented both quantitative and qualitative differences between expert and novice pianists in how their performances improve with practice. Most studies of music performance focus on one of two issues: systematic expressive nuances that provide perceptual cues to musical structure (e.g. Drake & Palmer, 1993; Palmer, 1989; Penel & Drake, 1998; Repp, 1992a,b; Shaffer et al., 1985; Sloboda, 1983), or production errors that provide cues to planning processes (Palmer & Drake, 1997; Palmer & van de Sande, 1993, 1995; Repp, 1996). Both of these lines of research focus on welllearned performances by highly-trained adult performers, who exhibit ®ne temporal control and make very few errors. In contrast, the main ®ndings of our study of novice performances indicate that: (1) beginners make many more errors than expert pianists, (2) beginners make different types of errors than experts, (3) beginners exhibit a lack of temporal control, and (4) beginners' ability to plan future events is limited. Therefore, the main theoretical issues that arise in the study of skilled performance are supplanted in novice performance by considerations of relativetiming/pitch accuracy trade-offs, temporal control, and planning constraints. An original goal of this study was to adapt traditional issues studied in skilled performance to novice performance by investigating both expressive nuances and production errors in children's piano performances. However, the highly variable nature of

novices' timing made identi®cation of expressive nuances dif®cult. In addition to the pitch errors described for expert performances, novices' performances exhibited many types of timing errors, including corrections, pauses, and 24 C. Drake, C. Palmer / Cognition 74 (2000) 1±32 duration errors, which are largely absent in adults' well-learned performances (Palmer & Drake, 1997; Palmer & van de Sande, 1995). For example, approximately 25% of the most skilled performers' errors exhibited a timing component in error, whereas more than 75% of the least skilled performers' errors exhibited a timing component in error. Thus, evaluation of novices' performance must include consideration of both which pitches are produced and when they are produced. We have reported a methodology in this study for examining both pitch and temporal errors as indicators of how performers maintain temporal control and plan for future events. In the next section we focus on these ®ndings in terms of the four issues of skill acquisition raised in the introduction. 6.1. Four issues of skill acquisition in music performance 6.1.1. Traditional measures of skill: speed and accuracy Both performance tempo and accuracy improved across skill levels (demonstrating long-term learning) and with practice (demonstrating short-term learning). More skilled pianists made fewer errors and performed at a faster rate. The observed initial rapid improvement in early practice trials and a gradual slowing in improvement with further practice is consistent with the power law described previously for many perceptual, cognitive, and motor tasks (Newell, 1980; Rosenbaum, 1991). Both the linear and logarithmic ®ts are within the range of those reported previously (here r 2 0:36±0.99 compared with r 2 0:40±0.99 in Newell & Rosenbloom (1981)). In general, a power function ®t the performance tempo and accuracy curves better than a linear function, demonstrating that the power function also applies for a complex task such as music performance. Most important, the power function held at initial stages in learning (the ®rst eleven practice trials, compared with hundreds or thousands of trials in previous studies), in contrast to views that a poor ®t for early trials re¯ects increased noise, associated with differences in learning strategies or prior knowledge (Ivry, 1996). 6.1.2. Additional characteristic of skill: ¯uency relative timing and pitch accuracy We did not observe a speed/accuracy trade-off in music performances at any skill level or point in practice, consistent with previous studies of skilled music performances (Palmer & van de Sande, 1993, 1995). This failure may be related to the additional temporal constraints imposed on music performance: in contrast to tasks such as typing, chess and problem solving, in which speed and accuracy are part of the task demands (MacKay, 1971), the task demand in music performance is not to play as fast as possible, but to respect the rhythmic structure speci®ed in the relative timing of sequence events. We report instead a relative-timing/pitch accuracy trade-off in both novice and expert piano performances for a novel musical sequence: pianists tended to concentrate either on performing the correct pitches or on performing at the correct moment in time. Least skilled pianists' performances exhibited high proportions of timing errors, whereas more skilled pianists exhibited mainly pitch errors. At early levels of skill acquisition, pianists tend to focus preferentially on producing the correct C. Drake, C. Palmer / Cognition 74 (2000) 1±32 25 pitches; only later are they able to master the temporal constraints, and thus successfully perform the relative timing and pitch accuracy required of the musical sequence. This relative-timing/pitch accuracy trade-off across skill levels was observed within skill level over practice trials as well; proportionally more timing errors relative to pitch errors were evidenced in initial trials, compared with later trials. Additional support for a relative timing/pitch accuracy trade-off is the frequency of joint pitch/duration errors. They occurred more often, and arose from the same sequence location more often, than expected from individual components. An absence of independence suggests that the pitch and time components intended for each event are accessed together rather than separately. Thus, dimensions of pitch and time interact at the level of individual events as well as at the level of entire sequences. It is possible that stimulus dif®culty, which

may have changed across groups, contributed to the timing/accuracy trade-off. Difficulty level was only partially controlled in this naturalistic design, and the higher skill levels exhibited lower error rates. However, comparisons of well-learned performances by novice and skilled pianists that contain equivalent error rates confirm that novices tend to produce temporally disruptive errors such as corrections, whereas advanced pianists produce pitch errors almost exclusively (Palmer & Drake, 1997). 6.1.3. Mastering the temporal structure of music: temporal continuity, underlying beat, and metrical structure Performers' abilities to produce the temporal structure of music can be decomposed into several specific timing skills. One skill is the ability to perform in a continuous manner without pausing or making corrections, which increased across skill levels. Pauses and other forms of temporal disruption have been examined as an index of mental load or production difficulty in complex tasks such as speech (Fodor et al., 1974). Another timing skill is the ability to respect the beat, which also increased across skill levels. A final timing skill is sensitivity to metrical structure, which all groups (except the least skilled) demonstrated: errors occurred more often than chance at the ends of metrical units (barlines) than at other locations, and this ability improved with skill level. These findings are consistent with both performance studies (Palmer, 1989; Palmer & Drake, 1997) and perceptual studies (Palmer & Krumhansl, 1990; Yee, Holleran & Jones, 1994) that demonstrate musicians' increased sensitivity to additional hierarchical levels of temporal structure. The ability to maintain temporal continuity improved with both short- and long-term learning, but improvements in the underlying beat and sensitivity to metrical structure have been demonstrated only over (long-term) learning. Findings that incorporate larger error data sets may also reveal short-term improvements. 6.1.4. Preparation of movements: the ability to plan and the range of planning Another hallmark of performance is the ability to prepare actions in advance of their production, or to plan. Future-oriented planning was indexed in music performance by the proportion of anticipatory errors (those reflecting influences of events intended for future events) relative to perseveratory errors (those reflecting the continuing influence of events that have already been performed). Our findings 26 C. Drake, C. Palmer / Cognition 74 (2000) 1±32 that anticipatory behavior increased both with skill level and with practice extend previous findings in music performance and speech production (Dell et al., 1997; Palmer & Drake, 1997) to indicate that experience with a task reflects more futureoriented behavior and less emphasis on the past. Anticipatory and perseveratory errors arise in associative models of planning in sentence production from activation rising more quickly for unintended sequence events than for intended (correct) elements, due to priming from other related events or nodes (Dell, 1986; MacKay, 1987). The Dell et al. (1997) model of sentence production predicts an increased anticipatory proportion of errors as practice enhances the activation of events intended for the present and future relative to the past. In addition, their model predicts an inverse relationship between anticipatory proportion and overall error rate, arising from a combination of practice, production rate, activation rate, and decay parameters. Our findings indicate that musicians were indeed more likely to anticipate as practice increased, and the anticipatory proportion was inversely related to overall error rates, both across skill levels and practice. The inverse relationship between error rates and anticipatory behavior, referred to as a general anticipatory effect, extends the speech error findings to a task that entails strict temporal constraints (Dell et al., 1997). These similarities suggest that serial order processes in sequence production reflect similar planning abilities at least across domains of music and speech. A related indicator of planning is the ability to plan for events from larger ranges of the sequence, referred to as range of planning. Errors tended to reflect events from increasingly larger ranges of the sequence at advanced skill levels, similar to previous findings with well-learned music performances (Palmer & Drake, 1997). In fact, the average number of events spanned by errors in performances of unfamiliar music (two to seven

events) was less than the typical range seen in welllearned performances (six to nine events) by pianists of similar skill levels (beginners, intermediates, and advanced pianists). Thus, this measure of planning scope suggests that the availability of elements from longer sequence ranges increases with both short-term practice with a piece and long-term experience. Furthermore, range of planning abilities are related to temporal control. There was an inverse relationship across skill level between performers' (pitch) range of planning and the frequency of temporal disruptions. These ®ndings suggest similar underlying cognitive abilities for temporal control and planning processes that develop with skill. 6.2. Integrating hallmarks of music performance: temporal control as an indicator of planning constraints Two noteworthy improvements in music performance over both short-term and long-term learning are highlighted in this study, namely, the progressive mastering of temporal constraints and increased planning abilities. The strong relation between these abilities suggests that the two phenomena are re¯ections of the same underlying cognitive processes. We ®rst identify functional characteristics of performance planning, and then suggest how breakdowns in temporal control re¯ect this planning. C. Drake, C. Palmer / Cognition 74 (2000) 1±32 27 6.2.1. The planning process Music performance in a sight-reading task involves the identi®cation of musical events, the conceptual interpretation of their structural relationships, and the coordination of movements necessary to produce them. Most studies of (adult) skilled performance evaluate the ®nal product: the often-memorized production of a welllearned musical piece. These studies usually indicate an error-free, smooth, uninterrupted musical ¯ow. In contrast, performances of novel musical pieces often lead to errorful performances that are temporally irregular (especially for novices). Such breakdowns in performance provide an opportunity to identify certain functional characteristics of the cognitive abilities underlying performance, under less-thanideal conditions. Based on these observations and related ®ndings, we identify the following functional characteristics: [2] Segmentation of a sequence. A long musical sequence is not planned all at once, but instead is broken down into short segments that contain several events (van Galen & Wing, 1984). Elements within a segment are more likely to be planned at the same time because they are simultaneously accessible; as a result, the relative timing of events within a segment tends to be respected, and events within a segment are more likely to interact with each other in errors. [2] Size of segments. Segment size is determined primarily by perceptual and memory constraints that limit the number of accessible events at a given time during the performance. However, segments may not reach the maximal planning capacity in sight-reading tasks because structural relationships (such as phrase boundaries) and physical relationships (such as note spacing) can lead to the creation of segment boundaries before the maximal planning capacity is reached (Sloboda, 1977). [2] Hierarchical structure. Planning over longer time-spans is facilitated if the sequence is conceived in a hierarchical fashion. Temporally remote, non-adjacent events that are related at higher hierarchical levels, such as metrical levels, as may be conceived and planned simultaneously. Relative timing can be respected over a longer time-spans than individual segments, and increasingly remote events become simultaneously accessible in a larger range of planning as hierarchical relationships are learned. [2] Planning near segment boundaries. Although planning occurs continuously during performance, planning for a future segment is concentrated near the end of the current segment. Indicators of breakdowns in planning (such as pauses or corrections) are more likely to occur at positions at which demands on planning are greatest: near the ends of segments. [2] Simultaneous planning and execution. Because music performance tasks contain strong temporal demands (not allowing a start-and-stop strategy), events must be executed at the same time as future events are being planned, within the available processing time (determined by temporal factors such as production rate and intended event duration). Breakdowns in performance are more likely to occur when there is insuf®cient processing

time. Certain predictions for skill acquisition processes arise from these cognitive 28 C. Drake, C. Palmer / Cognition 74 (2000) 1±32 abilities. First, skilled performers develop an increased range of planning, which arises from both increased planning capacities and heightened sensitivity to hierarchical structures relating non-adjacent sequence elements. Second, skilled individuals develop the ability to anticipate; this ability increases both with long-term and short-term learning. Third, requirements on processing time are reduced with increased skill, leading to greater ability to plan and execute simultaneously, as re¯ected in fewer temporal disruptions. Finally, increased segment size leads to fewer segment boundaries. Because temporal disruptions occur more frequently when demands of planning the next segment increase (near segment boundaries such as metrical boundaries), the number of opportunities for temporal disruptions decreases as the range of planning increases, leading to an inverse relationship. The ®ndings reported here are consistent with each of these predictions. 6.2.2. Relation between planning and temporal control Under the strict temporal constraints imposed in music performance, performers must produce the correct events (what) and produce them at the correct moment in time (when). Production errors, or breakdowns in planning, arise when performers cannot produce the intended events within the time available. When this happens, two possible scenarios arise: the ®rst is to focus on the event contents and correctly produce the intended pitches, at whatever time they become available. Another is to focus on the time at which the contents must be produced, and the content re¯ects whatever events are accessible at the correct time. The novice performances described in this study typify the ®rst solution; their performances are frequently interrupted by pauses, corrections, and duration errors. Additional processing time is obtained simply by stopping: future events are performed only after planning has been completed. Because segments are short (as evidenced by small ranges of planning), there are frequent opportunities for temporal breakdown at boundaries. The skilled performances typify the second solution. The correct pitches and relative timing can be planned ef®ciently in the available processing time because skilled performers require less processing time, and so performance is accurate in relation to both pitch and time. Moreover, planning segments are long, resulting in fewer opportunities for temporal disruption. Skilled performances of dif®cult tasks (such as sight-reading, or performing at a fast tempo) still re¯ect the second solution; inadequate processing time results in suboptimal planning, leading to temporally correct performances that contain more pitch errors. Thus, in skilled performance, breakdowns in planning are primarily in pitch, and rarely in time. Although the in¯uence of timing demands on planning constraints (and vice versa) may not come as a surprise to researchers of music performance, considerations of timing do not enter into many theories of skilled behaviors. This may be attributable in part to the fact that most research examines well-learned behaviors, or behaviors of highly skilled individuals, who may exhibit very little temporal disruption. However, the greatest cognitive change in sequence learning may be found at relatively inexperienced skill levels, and so novice performance may be an important indicator of the cognitive bases of skill acquisition (Adams, 1987; Palmer & C. Drake, C. Palmer / Cognition 74 (2000) 1±32 29 Drake, 1997; Palmer & Meyer, in press). That ®nding, combined with the parallels reported here in planning behaviors that transcend music and speech, suggest that examination of sequence timing may explicate planning abilities that underlie many complex skills.

References Adams, J. A. (1987). Historical review and appraisal of research on the learning, retention, and transfer of human motor skills. Psychological Bulletin, 101, 41±74. Anderson, J. R. (1982). Acquisition of cognitive skills. Psychological Review, 89, 369±406. Bamberger, J. (1980). Cognitive

structures in the apprehension and description of simple rhythms. Archives de Psychologie, 48, 171±199. Chi, M.T.H., Glaser, R., Farr, M.J. (1988). The nature of expertise. Hillsdale, NJ: Erlbaum. Clarke, E. F. (1985). Structure and expression in rhythmic performance. In P. Howell, I. Cross, & R. West, Musical structure and cognition, (pp. 209±236). London: Academic Press. Cooper, G., & Meyer, L. B. (1960). The rhythmic structure of music, Chicago: The University of Chicago Press. Crossman, E. R. F. W. (1959). A theory of the acquisition of speed skill. Ergonomics, 2, 153±166. Dell, G. S., Burger, L. K., & Svec, W. R. (1997). Language production and serial order: a functional analysis and a model. Psychological Review, 104 (1), 123±147. Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. Psychological Review, 93 (3), 283±321. Dowling, W. J., & Harwood, D. L. (1986). Music cognition, Orlando: Academic Press. Drake, C., & GeÂrard, C. (1989). A psychological pulse train: how young children use this cognitive framework to structure simple rhythms. Psychological Research, 51, 16±22. Drake, C., & Palmer, C. (1993). Accent structures in music performance. Music Perception, 10 (3), 343± 378. Drake, C., Dowling, J., & Palmer, C. (1991). Accent structures in the reproduction of simple tunes by children and adult pianists. Music Perception, 8 (3), 313±332. Drake, C. (1993a). Reproduction of musical rhythms by children, adult musicians and adult non-musicians. Perception and Psychophysics, 53 (1), 25±33. Drake, C. (1993b). In¯uence of age and experience on timing and intensity variations in the reproduction of short musical rhythms. Psychologica Belgica, 33-2, 217±228. Drake, C. (1997). Motor and perceptually preferred synchronisation by children and adults: binary and ternary ratios. Polish Quarterly of Developmental Psychology, 3 (1), 41 ±59. Drake, C. (1998). Psychological processes involved in the temporal organization of complex auditory sequences: universal and acquired processes. Music Perception, 16 (1), 11±26. Ericsson, K. A., & Staszewski, J. (1989). Skilled memory and expertise: mechanisms of exceptional performance. In D. Klahr, & K. Kotovsky, Complex information processing: the impact of Herbert A. Simon, (pp. 235±367). Hillsdale, NJ: Erlbaum. Ericsson, K. A., & Smith, J. (1991). Prospects and limits of the empirical study of expertise: an introduction. In K. A. Ericsson, & J. Smith, Toward a general theory of: prospects and limits, (pp. 1±39). Cambridge: Cambridge University Press. 30 C. Drake, C. Palmer / Cognition 74 (2000) 1±32 Ericsson, K. A., Krampe, R., & Tesch-Romer, C. (1993). The role of deliberate practice in the acquisition of expert performance. Psychological Review, 100, 363±406. Essens, P. J., & Povel, D. J. (1985). Metrical and non-metrical representations of temporal patterns. Perception and Psychophysics, 37 (1), 1±7. Essens, P. J. (1986). Hierarchical organization of temporal patterns. Perception and Psychophysics, 40, 69±73. Essens, P. J. (1995). Structuring temporal sequences: comparison of models and factors of complexity. Perception and Psychophysics, 57 (4), 519±532. Fodor, J. A., Bever, T. G., & Garrett, M. F. (1974). The psychology of language: an introduction to psycholiguistics and generative grammar, New York: McGraw-Hill. Garcia-Albea, J. E., del Viso, S., & Igoa, J. M. (1989). Movement errors and levels of processing in sentence production. Journal of Psycholinguistic Research, 18, 145±161. Garrett, M. F. (1975). The analysis of sentence production. In G. H. Bower, Psychology of learning and motivation, (pp. 133±177). New York: Academic Press. Garrett, M. F. (1980). Levels of processing in sentence production. In B. Butterworth, Language production: Speech and Talk, (pp. 177±220). London: Academic Press. Gruson, L. M. (1988). Rehearsal skill and musical competence: does practice make perfect? In J. A. Sloboda, Generative processes in music: the psychology of performance, improvization, and composition, (pp. 91±112). New York: Oxford University Press. Ivry, R. (1996). Representational issues in motor learning: phenomena and theory. In H. Heuer, & S. W. Keele, Handbook of perception and action, vol 2: motor skills, (pp. 263±330). San Diego, CA: Academic Press. Keele, G. E., & Summers, J. (1976). The structure of motor programs. In Stelmach, Motor control: issues and trends, (pp.

109±142). New York: Academic Press. Kolers, P. A. (1975). Memorial consequences of automatized encoding. Journal of Experimental Psychology: Human Learning and Memory, 1 (6), 689±701. Lehmann, A. C., & Ericsson, K. A. (1993). Sight-reading ability of expert pianists in the context of piano accompanying. Psychomusicology, 12, 182±195. Lerdahl, F., & Jackendoff, R. (1983). A generative theory of tonal music, Cambridge Massachusetts: MIT Press. Logan, G. D. (1988). Toward an instance theory of automatization. Psychological Review, 95, 492± 527. MacKay, D. (1971). Stress pre-entry in motor systems. American Journal of Psychology, 84, 31±35. MacKay, D. G. (1982). The problems of ¯exibility, ¯uency, and speed-accuracy trade-off in skilled behavior. Psychologicul Review, 89, 483±506. MacKay, D. G. (1987). The organization of perception and action: a theory for language and other cognitive skills, New York: Springer. McKenzie, C. L., Nelson-Schultz, J. A., & Wills, B. L. (1983). A preliminary investigation of motor programming in piano performance as a function of skill level. In D. R. Rogers, & J. A. Sloboda, The acquisition of symbolic skills, (pp. 283±292). New York: Plenum Press. Newell, A. (1980). Harpy, production systems and human cognition. In R. Cole, Perception and production of ¯uent speech, (pp. 289±380). Hillsdale, N.J: Erlbaum. Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson, Cognitive skills and their acquisition, (pp. 1±55). Hillsdale, NJ: Erlbaum. Palmer, C., & Drake, C. (1997). Monitoring and planning capacities in the acquisition of music performance skills. Canadian Journal of Experimental Psychology, 51 (4), 369±384. Palmer, C., & Krumhansl, C. L. (1990). Mental representations for musical meter. Journal of Experimental Psychology: Human Perception and Performance, 16 (4), 728±741. Palmer, C., & Meyer, R. K. (in press). Conceptual and motor learning in music performance. Psychological Science,. Palmer, C., & van de Sande, C. (1993). Units of knowledge in music performance. Journal of Experimental Psychology: Learning, Memory, and Cognition, 19 (2), 457±470. Palmer, C., & van de Sande, C. (1995). Range of planning in skilled music performance. Journal of Experimental Psychology: Human Perception and Performance, 21, 947±962. C. Drake, C. Palmer / Cognition 74 (2000) 1±32 31 Palmer, C. (1989). Mapping musical thought to musical performance. Journal of Experimental Psychology: Human Perception and Performance, 15 (12), 331±346. Palmer, C. (1997). Music Performance. Annual Review of Psychology, 48, 115±138. Parncutt, R. (1993). A perceptual model of pulse salience and metrical accent in musical rhythms. Music Perception, 11 (4), 409±464. Penel, A., & Drake, C. (1998). Sources of timing variations in music performance: a psychological segmentation model. Psychological Research, 61 (1), 12±32. Povel, D. J. (1981). Internal representation of simple temporal patterns. Journal of Experimental Psychology: Human Perception and Cognition, 7, 3±18. Repp, B. H. (1992). Probing the cognitive representation of musical time: structural constraints on the perception of timing perturbations. Cognition, 44, 241±281. Repp, B. H. (1992). Diversity and commonality in music performance: an analysis of timing microstructure in Schumann's TraÈumerei. Journal of the Acoustical Society of America, 92, 2546±2568. Repp, B. H. (1996). The art of inaccuracy: why pianists' errors are dif®cult to hear. Music Perception, 14, 161±183. Rosenbaum, D. A. (1991). Human motor control, San Diego: Academic Press. Schneider, W. S. R. (1977). Controlled and automatic human information processing: I. detection, search, and attention. Psychological Review, 84 (1), 1±66. Shaffer, L. H. (1980). Analysing piano performance: a study of concert pianists. In G. E. Stelmach, & J. Requin, Tutorials on motor behavior, (pp. 443±455). Amsterdam: North-Holland. Shaffer, L. H., Clarke, E. F., & Todd, N. P. (1985). Metre and rhythm in piano playing. Cognition, 20, 61± 77. Sloboda, J. A. (1977). Phrase units as determinants of visual processing in music reading. British Journal of Psychology, 68, 117±124. Sloboda, J. A. (1983). The communication of musical metre in piano performance. Quarterly Journal of Experimental Psychology, 35, 377±396. Smith, J. (1983). Reproduction and representation of musical rhythm: the effects of

musical skill. In D. R. Rogers, & J. A. Sloboda, Acquisition of symbolic skill, (pp. 273±283). New York: Plenum. van Galen, G., & Wing, A. M. (1984). The sequencing of movements. In M. Smyth, & A. M. Wing, The psychology of human movement, (pp. 153±182). London: Academic Press. Yee, W., Holleran, S., & Jones, M. (1994). Sensitivity to event timing in regular and irregular sequence: in¯uences of musical skill. Perception and Psychophysics, 56, 461±471. Yeston, M. (1976). The strati®cation of musical rhythm, New Haven: Yale University Press. 32 C. Drake, C. Palmer / Cognition 74 (2000) 1±32

Abstract

Memories are easier to relearn than learn from scratch. This advantage, known as savings, has been widely assumed to result from the reemergence of stable long-term memories. In fact, the presence of savings has often been used as a marker for whether a memory has been consolidated. However, recent findings have demonstrated that motor learning rates can be systematically controlled, providing a mechanistic alternative to the reemergence of a stable long-term memory. Moreover, recent work has reported conflicting results about whether implicit contributions to savings in motor learning are present, absent, or inverted, suggesting a limited understanding of the underlying mechanisms. To elucidate these mechanisms, we investigate the relationship between savings and long-term memory by experimentally dissecting the underlying memories based on short-term (60-s) temporal persistence. Components of motor memory that are temporally-persistent at 60 s might go on to contribute to stable, consolidated long-term memory, whereas temporally-volatile components that have already decayed away by 60 s cannot. Surprisingly, we find that temporally-volatile implicit learning leads to savings, whereas temporally-persistent learning does not, but that temporally-persistent learning leads to long-term memory at 24 h, whereas temporally-volatile learning does not. This double dissociation between the mechanisms for savings and long-term memory formation challenges widespread assumptions about the connection between savings and memory consolidation. Moreover, we find that temporally-persistent implicit learning not only fails to contribute to savings, but also that it produces an opposite, anti-savings effect, and that the interplay between this temporally-persistent anti-savings and temporally-volatile savings provides an explanation for several seemingly conflicting recent reports about whether implicit contributions to savings are present, absent, or inverted. Finally, the learning curves we observed for the acquisition of temporally-volatile and temporally-persistent implicit memories demonstrate the coexistence of implicit memories with distinct time courses, challenging the assertion that models of context-based learning and estimation should supplant models of adaptive processes with different learning rates. Together, these findings provide new insight into the mechanisms for savings and long-term memory formation.

Figures

Data Availability: Individual quantitative observations that underlie the data summarized in the figures and results of this paper can be found in Zenodo (doi: 10.5281/zenodo.7668262) Extended data and analysis code are maintained on a GitHub repository: https://github.com/AlkisMH/Savings_vs_Long_Term_Memory.

Competing interests: The authors have declared that no competing interests exist.

Abbreviations: CUHS, Committee on the Use of Human Subjects; IQR, interquartile range; UDL, use-dependent learning; VMR, visuomotor rotation

Introduction

Memories, both declarative and procedural, are easier to relearn than to learn from scratch. This advantage, known as savings, was first appreciated in Hermann Ebbinghaus's seminal work [1], in which he observed that relearning a forgotten list of words was faster than learning a novel list. Savings has

since been demonstrated in a plethora of different paradigms, including cognitive tasks in humans [2,3], operant conditioning in animals [4–6], and motor tasks in humans such as saccade adaptation [7], force-field adaptation [8–11], visuomotor adaptation [12–21], and gait adaptation [22,23].

Previous research has generally maintained that savings results from the recall of a previously consolidated long-term memory [14,24,25]. The ability to form enduring long-term memories is one of the most remarkable biological abilities, with some memories lasting a lifetime despite incessant neural plasticity. Elderly Danes recalled mundane details—such as the weather—surrounding the news of their country's invasion and liberation during WWII more than 50 years later [26], and similar examples of memories surrounding shocking and/or consequential events abound [27–29]. People recognize high school classmates decades later [30], tacitly use synthetic grammar rules years after training [31], and identify specific images months [32] or even years [33] later. Long-term memories emerge through consolidation, a process driven by molecular mechanisms that facilitate synaptic activity [34,35] and can be mediated through the hippocampus [36,37]. In motor tasks, the presence or absence of savings itself has often been taken as a litmus test for whether a previously trained memory has been consolidated, at timescales ranging from hours to months [14,24,25]. Consequently, in terms of underlying mechanisms, savings has been further suggested to result from the following: (1) the unmasking of a slower-learning, strong-retention process in a multi-rate learning model [10]; (2) context- or relevance-based switching between such multiple slow processes, each specific to a different memory [38–41]; or (3) reverting to the memory of a previously learned motor plan that was reinforced by success or mere repetition [16,42]. All of these proposed mechanisms focus on savings as the manifestation of a latent, stable, consolidated motor memory that is robust to both interference and the passage of time.

In line with this idea, a recent but influential view has proposed that savings in motor adaptation specifically results from the recall of an explicit strategy [18,43,44], whereas the implicit component of visuomotor learning does not contribute to savings [44]. These studies provided clear evidence for explicit savings, but the paradigms they used elicited little implicit adaptation, limiting the power to assess implicit savings. More recent studies, which elicited greater implicit adaptation, have led to disparate findings, concluding that implicit adaptation is either faster during relearning [19,20] or slower [45]. The Avraham [45] and Albert [20] studies are of particular interest, as they were both designed to isolate implicit adaptation (albeit using different paradigms) and yet reached opposite conclusions.

What could explain this apparent discrepancy? A possibility is that implicit adaptation may not be monolithic. It may consist of distinct components that

are relearned at different rates and differentially elicited in these different paradigms. In fact, an intriguing alternative to recall-based mechanisms is that savings arises from changes in learning rate [11,15,23], an idea reinforced by recent work which has demonstrated that the rate at which learning occurs is systematically modulated by specific characteristics of the learning environment. These characteristics include the amount of task-relevant variability present before learning [46,47], the balance between sensory uncertainty and uncertainty about state estimation [48,49], prior exposure to perturbations characterized by a similar covariance structure between learning parameters [50,51], prior exposure to similar motor errors [21,52], and the trial-to-trial consistency of the learning environment [53,54]. In particular, high-consistency environments, whereby perturbations tend to persist from one trial to the next and thus confer more predictability to the imposed perturbation, can strongly increase learning rates (up to 3×). This is a critical finding as far as the study of savings is concerned: The adaptation paradigms used to study savings usually consist of the same perturbation being active for a large number of trials (usually 60 to 120), resulting in highly consistent errors, which would in turn lead to strongly increased learning rates [52–55]. This mechanism could lead to savings by enabling the faster relearning of a short-term memory, as opposed to the reemergence of a long-term, stable memory. This is consistent with recent results indicating that prior experience with high-consistency errors, as opposed to the repetition of successful actions, leads to savings [21].

Here, we compare the mechanisms that lead to savings and those that lead to the formation of stable, long-term motor memories. We hypothesized that dissecting motor adaptation into specific memory components based on temporal persistence could shed light into these mechanisms. We use short 60-s time delays to dissect overall adaptation into 2 components: temporally-volatile adaptation, which would decay during this time delay, and temporally-persistent adaptation, which would survive the delay [56–59]. Since it decays away in just 60 s, temporally-volatile adaptation would not lead to long-term memory that is associated with stability across timescales that are orders of magnitude greater [9,25,60]. Hence, mechanisms leading to long-term memory would be contained within temporally-persistent adaptation, and thus, consolidation-dependent savings would predict faster learning solely for temporally-persistent adaptation. Conversely, if we found faster relearning solely for temporally-volatile adaptation, that would indicate a savings mechanism that is not driven by long-term memory.

Remarkably, we find that savings is driven not by the reemergence of temporally-persistent motor memories, but instead by faster relearning of temporally-volatile memories. We go on to find that these temporally-volatile memories responsible for savings in our paradigm represent implicit, rather

than explicit, adaptation. When we measure the long-term retention of the temporally-persistent and temporally-volatile components, however, we find that it is temporally-persistent adaptation, not temporally-volatile adaptation, that leads to long-term memory. Together, these findings demonstrate a clear double dissociation between savings and long-term memory.

Results

We designed a set of experiments to elucidate the mechanisms for savings and long-term memory and investigate the relationship between them. We began by investigating whether savings, the faster relearning of a previously learned adaptation, is driven by the reemergence of a previously consolidated temporally-persistent memory, or by a propensity for faster acquisition of a transient, temporally-volatile memory. In particular, we created a paradigm to dissect initial adaptation, the washout of adaptation, and savings in readaptation into temporally-persistent and temporally-volatile components. We first investigated the dynamics by which temporally-persistent and temporally-volatile memories decay during a washout period following initial adaptation, as savings can arise from the incomplete washout of a component of adaptation [10,61]. This allowed us to compare the rates of unlearning for temporally-persistent and temporally-volatile adaptation during washout, and critically, to measure the initial value of both temporally-persistent and temporally-volatile memories prior to readaptation, so that savings could be accurately assessed for both. We next examined how savings depends on temporally-persistent and temporally-volatile memories by measuring savings separately for these 2 components of motor adaptation, allowing us to determine whether one of these memories is specifically responsible for savings. We then investigated whether long-term memory, measured as the retention of a previously trained adaptation 24 h later, is associated with the temporally-persistent or the temporally-volatile component of adaptation.

Measuring temporally-volatile and temporally-persistent contributions to savings

For Experiments 1 and 2, we recruited N = 40 subjects and trained them on a 30° visuomotor rotation (VMR) [12,13,15–18,62–64] (**Fig 1A and 1B**). After 80 trials of initial training, subjects were tested for savings after either a short (40-trial) washout period, which was previously reported to be sufficient for the washout of overall VMR adaptation [15], or a longer (800-trial) washout period we employed to effect a more definitive washout. We also used the data from this 800-trial washout to trace out the time course of unlearning for both the temporally-persistent and temporally-volatile components of adaptation; this unlearning would encompass both active unlearning (i.e., relearning the baseline behavior) and natural trial-to-trial decay of adaptation [17,65–68].

Each subject experienced both types of washout duration following training (see **Fig 1C**). In Experiment 1 ($N = 20$), the short washout period was presented first and the long washout period second. In Experiment 2 ($N = 20$), this order was flipped (**Fig 1C**, for a detailed description see Materials and methods).

Fig 1. Experiment setup and training paradigm.

(a) Experiment setup. Subjects made point-to-point reaching movements on a digitizing tablet and received continuous visual feedback on a screen mounted above it. **(b)** VMR training. During baseline (left), the cursor follows the hand motion, whereas during training, cursor motion is skewed by 30° from the hand motion (in this example, counter-clockwise), resulting in a 30° error before adaptation (middle). If full adaptation is achieved, hand motion must completely counter the imposed rotation, corresponding to a 30° clockwise hand motion in this example (right). **(c)** Top: experiment schedule and raw data for Experiment 1. There were 3 phases: a baseline period followed by the initial 80-trial VMR training (average adaptation level shown in gray); a short, 40-trial washout period followed by retraining (green); and a long, 800-trial washout period followed by another 80-trial retraining session (blue). Red dashed vertical lines indicate trials conducted after 60-s delays to isolate temporally-persistent adaptation. Brown dashed vertical lines indicate trials following rest breaks. Note that, during the washout periods, adaptation peaks on these delay and rest break trials, illustrating a slower washout for temporally-persistent vs. temporally-volatile adaptation. Bottom: same but for Experiment 2, where the long, 800-trial washout period came first. Error bars indicate SEM. Underlying data supporting this panel can be found in file Exp_1_2_data.mat. **(d)** Dissection of adaptation into temporally-persistent and temporally-volatile components. Throughout the experiment, we used 60-s delays to allow temporally-volatile adaptation to decay. The amount of adaptation on the trial following such a delay (open gray circle) was taken as a measure of temporally-persistent adaptation, whereas the average amount of adaptation 2 trials before and 2 trials after this post-delay trial (filled gray circles) was taken as a measure of overall adaptation. Temporally-volatile adaptation was operationally defined as the difference between overall and persistent adaptation.

https://doi.org/10.1371/journal.pbio.3001799.g001

Throughout these experiments—during both learning and washout—we occasionally inserted 1-min delays which would allow for temporally-volatile adaptation to decay. Since the 1-min delays we imposed amount to 2.5 to 4× the time constant for decay of temporally-volatile adaptation [56–58], approximately 95% decay of volatile adaptation would be expected, effectively isolating temporally-persistent adaptation. We operationally defined temporally-persistent adaptation as the adaptation measured during the post-delay trial and temporally-volatile adaptation as the difference between overall adaptation (itself taken as the average adaptation in the 2 preceding and 2 proceeding non-delay trials) and temporally-persistent adaptation (**Fig 1D**, also see Materials and methods).

Temporally-persistent adaptation washes out more slowly than overall adaptation

The data from the long, 800-trial washout period allowed us to carefully examine the time course of unlearning for both the overall adaptation and for the temporally-persistent component of it. Analysis of the washout curves revealed that overall adaptation displayed rapid unlearning; however, persistent adaptation (circles in **Fig 2A**) was unlearned much more slowly. We found that by trials 16 to 25, labeled as "early washout" in **Fig 2A**, overall adaptation had already dropped below 10% of the pre-washout asymptotic adaptation level, whereas about 40% of pre-washout persistent learning remained. By trials 51 to 150, labeled as "mid washout," overall adaptation had dropped below 3%, whereas about 20% of persistent adaptation still remained (**Fig 2A**, see inset). Correspondingly, we found the retention of persistent learning to be significantly greater than overall learning in both early washout (t(23) = 4.8, $p = 6.9 \times 10^{-5}$ and mid washout periods (t(39) = 8.5, $p = 1.9 \times 10^{-10}$). To quantify the rate of unlearning during washout for both overall and persistent adaptation, we fit single exponential decay functions to the washout data (see Materials and methods). This revealed the time constants for unlearning to be 6-fold slower for temporally-persistent adaptation than for overall adaptation (median time constant estimated using bootstrap: 106.0 trials, interquartile range (IQR) [92.6 to 121.9] versus 17.4 trials, IQR [15.1 to 20.1], $p < 10^{-4}$, **Fig 2A**), in line with the higher retention we observed in the early- and mid-washout data.

original image**Fig 2. Temporally-persistent adaptation washes out more slowly than overall adaptation.**
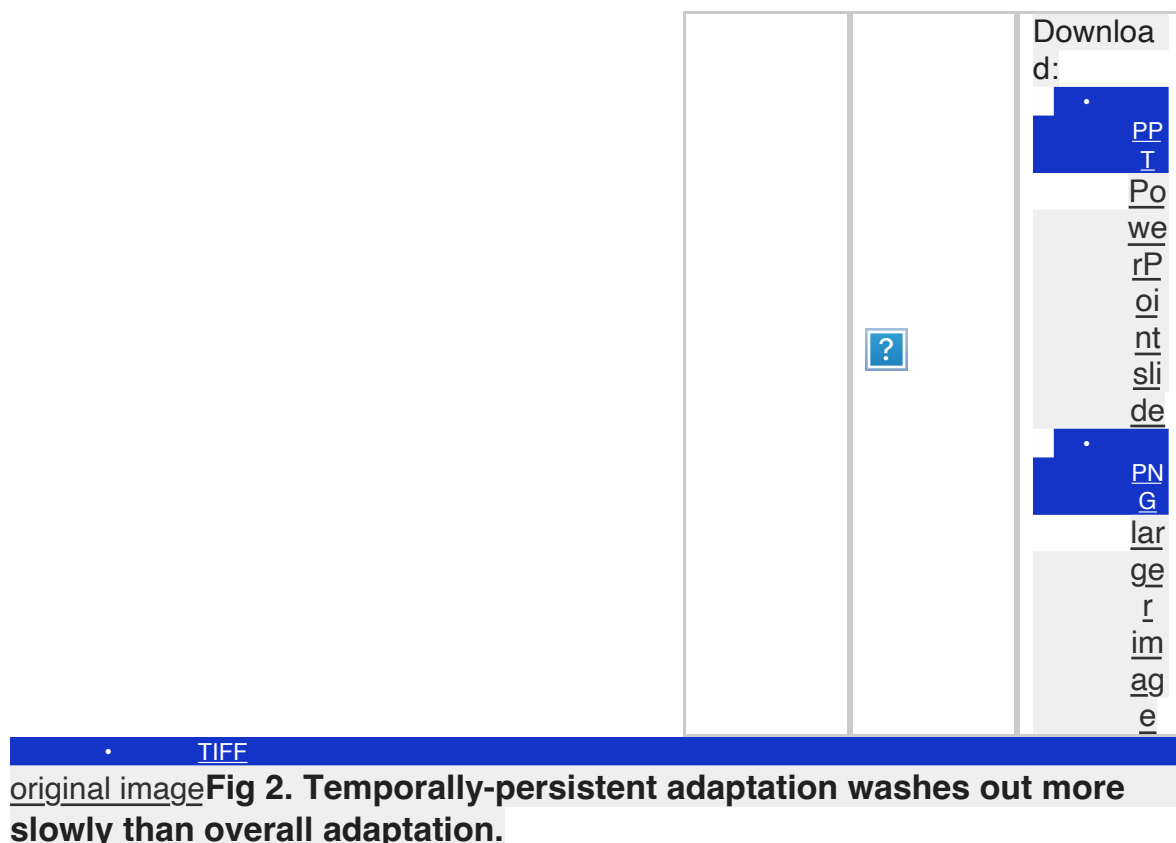
(a) Washout curves for the overall adaptation (shading indicates mean ± SEM) and temporally-persistent adaptation (circles) for both Experiment 1 (blue) and Experiment 2 (light blue), illustrating the contrast between rapid washout for overall adaptation and slower washout for temporally-persistent adaptation. The thick dashed or dotted lines indicate exponential fits. Inset: Three different measures of washout for overall vs. persistent adaptation. Retention expressed as a percentage of asymptote adaptation after 16–25 trials (left) or 51–150 trials (center) indicates slower washout for temporally-persistent compared to overall adaptation. Time constants for the washout curves (right) also show slower washout for temporally-persistent adaptation. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. **(b)** Residual adaptation before initial training (gray, left bar in each cluster), at the end of the 40-trial washout period (green, middle bar) and at the end of the 800-trial washout period (blue, right bar). The data show that the 40-trial washout period leaves a significant amount of temporally-persistent adaptation and a smaller but also significant amount of overall adaptation. Consequently, retraining after only 40 washout trials starts from a nonzero baseline. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. Underlying data supporting this figure can be found in file Exp_1_2_data.mat.

https://doi.org/10.1371/journal.pbio.3001799.g002

As a minor point, we also noticed that unlearning curves for temporally-persistent adaptation display somewhat greater retention during the early and mid-washout in Experiment 1 compared to Experiment 2 (**Fig 2A**). This might

reflect the difference in the amount of training between the 2 conditions, as 2 training blocks (160 trials in total) preceded this washout period in Experiment 1, whereas only a single training block (80 trials) preceded this washout in Experiment 2, due to the condition balancing (see **Fig 1**). In summary, we found that temporally-persistent adaptation is unlearned at a considerably slower rate than overall adaptation.

Residual adaptation prior to the onset of retraining

One consequence of the slower unlearning of persistent compared to overall adaptation is that, while the washout of overall adaptation can appear complete after a short 40 to 100 trial per direction washout period [15,16,18,24], substantial temporally-persistent adaptation can nevertheless remain. This suggests that longer washout periods may be required to examine savings independent of the effect of residual temporally-persistent adaptation and that measuring this residual adaptation prior to relearning may facilitate a better understanding of relearning behavior.

When we measured the residual overall and persistent adaptation before the onset of retraining, we found significant levels of both overall and temporally-persistent adaptation for the 40-trial washout but no significant residuals of the previous adaptation following the 800-trial washout period. In particular, we found small but significant residuals for overall adaptation at the end of the 40-trial washout periods, around 5% of pre-washout levels ($1.34 \pm 0.37°$, $t(39) = 3.6$, $p = 0.00087$ for Experiments 1 and 2 combined, with positive values indicating adaptation in the direction of the previously imposed VMR, **Fig 2B**). Note that these residuals were related to previous adaptation rather than a movement direction bias because both Experiments 1 and 2 were balanced, with 10 participants trained with clockwise, and 10 with counter-clockwise VMRs for each experiment. The residuals were even larger for temporally-persistent adaptation, in line with the substantially slower unlearning of temporally-persistent adaptation compared to overall adaptation we observed. In particular, we found that the residual persistent adaptation before the end of the 40-trial washout was around 25% of pre-washout persistent adaptation ($4.91 \pm 0.49°$ for Experiments 1 and 2 combined, $t(39) = 10.0$, $p = 2.7 \times 10^{-12}$). In contrast, the 800-trial washout period was sufficient to bring both overall adaptation and temporally-persistent adaptation back to baseline, with measured residuals of only 0% to 2% of pre-washout levels on average. These residuals were not consistently in the direction of the pre-washout adaptation and were not statistically significant. Overall adaptation at the end of the 800-trial washout was $-0.00 \pm 0.18°$ ($t(39) = -0.01$, $p = 0.9955$), whereas temporally-persistent adaptation was $0.13 \pm 0.32°$ ($t(39) = 0.44$, $p = 0.6772$), as shown in **Fig 2B**.

These results show that a prolonged washout period is required to eliminate residual temporally-persistent adaptation. As washout periods in previous experimental work on savings [15,16,18,24,64] are typically much shorter than the 800-trial washout period we examined, it is likely that the savings observed in these studies is, at least in part, driven by interactions between different components of adaptation that were not fully washed out prior to retraining—apparent savings—as suggested in Smith and colleagues [10]. In order to examine faster relearning that is not contaminated by such interactions—that is, examine true savings—one should ideally eliminate residual levels of overall, temporally-persistent, and temporally-volatile adaptation, or, at least, take these residual levels into account.

Savings is present even when previous adaptation is completely washed out

To investigate savings for overall and temporally-persistent adaptation, we compared the learning curves for retraining and initial training, shown in **Fig 3A and 3B** (gray: initial training; green: retraining after 40 washout trials; blue: retraining after 800 washout trials). We found that the adaptation levels achieved in the early adaptation period (trials 8 to 12 after perturbation onset, excluding trial 10 which was after a time delay), when learning was most rapid, were noticeably higher for relearning (24.6 ± 0.7° overall; 24.9 ± 0.6° and 24.4 ± 1.0° after the 40-trial and 800-trial washout periods, green and blue lines, respectively, with data combined across Experiments 1 and 2 in all cases) compared to initial training (18.8 ± 1.2°, gray; $p < 10^{-5}$ compared to relearning after either washout period or both periods combined). Because pretraining adaptation levels were not identical across conditions as shown in **Fig 2B** (1.34 ± 0.37°, −0.00 ± 0.18°, and −0.04 ± 0.24°, after short washout, long washout, and before initial adaptation), we normalized data to quantitatively compare learning and relearning curves independent of the effect of this residual pretraining adaptation. Specifically, we subtracted the pretraining adaptation separately for overall and temporally-persistent adaptation, and normalized each baseline-subtracted learning curve by the distance between baseline and the ideal adaptation level (Eq 2, see Materials and methods). These normalized data, plotted in **Fig 3C and 3D** express adaptation levels as a percentage of that required for full adaptation. In particular, normalized early adaptation (trial 10 after perturbation onset) was faster compared to initial adaptation (initial adaptation: 62.8 ± 4.1% versus relearning: 81.7 ± 2.5%; 40-trial and 800-trial washout data separately: 82.1 ± 2.1% and 81.4 ± 3.2%, respectively). We defined savings simply as the difference between these normalized adaptation data for the retraining versus the initial learning conditions. The top panel in **Fig 3E and 3F** shows an estimate of this savings measure in the overall adaptation. We find statistically significant savings for early adaptation (trial 10; savings of 18.9 ± 3.3% of the ideal adaptation, t(39) = 5.8, $p = 5.4 \times 10^{-7}$ [40-trial washout data: 19.3 ±

3.3%, t(39) = 5.8, p = 4.4 × 10^{-7}; 800-trial washout data: 18.6 ± 3.6%, t(39) = 5.1, p = 4.1 × 10^{-6}). Inspection of the overall adaptation data in the top panels of **Fig 3C and 3D** reveals that both the readaptation and initial adaptation curves asymptote near the ideal adaptation level, meaning that the room for improvement, and thus the capacity for savings, is reduced as training proceeds. In line with this observation, the savings we observed for mid (trial 40) and late (trial 70) overall adaptation were smaller than the savings observed for early (trial 10) adaptation (<10% of the ideal adaptation in all cases). The overall savings we observed at trials 40 and 70 were, however, statistically significant (t(39) = 2.6, p = 0.0063 for trial 40 and t(39) = 3.3, p = 0.0011 for trial 70), as shown in **Fig 3E and 3F**.
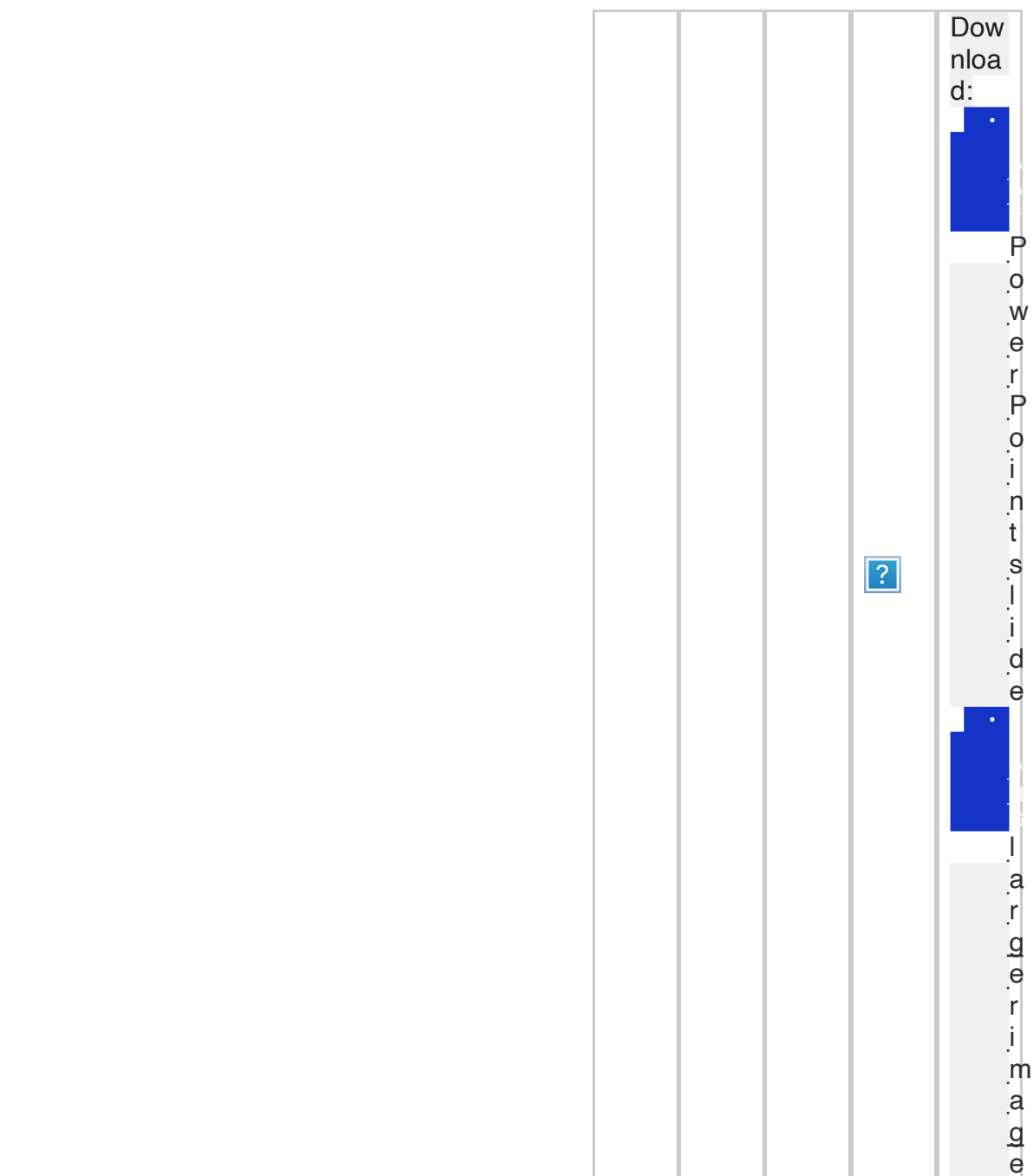
Fig 3. Temporally-volatile adaptation displays savings, but temporally-persistent adaptation does not.

(a, b) Learning curves for Experiments 1 and 2, showing overall and persistent adaptation (solid lines and circles, respectively). Persistent adaptation during training or retraining was assessed after a 60-s delay as shown in Fig 1. The dashed black line indicates ideal performance and error bars indicate SEM. Overall adaptation (solid lines) is faster for retraining after both a 40-trial washout period (green) and an 800-trial washout period (blue) when compared to initial adaptation (gray). In contrast, persistent adaptation (circles) is not faster during retraining than initial training. Note that

performance during the retraining period following a 40-trial washout (green) can be inflated because this washout is often incomplete for both overall and persistent learning, as indicated by nonzero pretraining levels as also shown in **2b**. Washout for both overall and persistent adaptation is complete for the 800-trial data (blue) in both Experiments 1 and 2 resulting pretraining levels that are essentially zero in all cases. Vertical red dashed lines indicate trials following 1-min delays. **(c, d)** Normalized overall, persistent, and volatile adaptation components in Experiments 1 and Experiment 2. Upper panels show the normalized learning curves for overall and temporally-persistent adaptation. Here, the raw data shown in A were normalized by subtracting out pretraining adaptation levels so that training-related changes in performance can be directly assessed and scaling ideal performance to 100% (see [Materials and methods](#), [Eq 2](#)). Whereas the overall adaptation is faster for retraining after both washout periods, persistent adaptation is not. Lower panels show the normalized learning curves for temporally-volatile adaptation. Temporally-volatile readaptation is consistently higher than initial adaptation, especially for the early (trial 10) time point, where savings should be most pronounced and where readaptation is between 2–3 times higher. **(e, f)** Savings for overall, temporally-persistent, and temporally-volatile adaptation in Experiments 1 and 2, defined as the difference between the normalized adaptation metrics displayed in panels **c and d** for relearning vs. initial learning (see [Materials and methods](#), [Eq 3](#)). Positive values for savings indicate faster relearning. We find substantial and significant savings especially early during training (trial 10) for both overall and temporally-volatile adaptation, but not for temporally-persistent adaptation, suggesting that savings is driven by temporally-volatile adaptation. Underlying data supporting this figure can be found in file Exp_1_2_data.mat.

Savings arises from the faster acquisition of temporally-volatile memories

Remarkably, we found that the savings observed in overall adaptation was due to temporally-volatile learning. We calculated savings for temporally-volatile adaptation (bottom row in **Fig 3E and 3F**) based on normalized volatile adaptation (bottom row of **Fig 3C and 3D**), which was computed as the difference between the normalized overall and normalized persistent adaptation (top row of **Fig 3C and 3D**). We found that volatile adaptation during early training (trial 10) was 2- to 3-fold faster for retraining than for initial training after both the short and long washout periods in both experiments (as shown in the bottom row of **Fig 3C and 3D**). Specifically, we found that trial 10 volatile readaptation was 52.2 ± 3.8% of the ideal adaptation to the 30° VMR with data after both types of washout combined versus 22.0 ± 4.2% for initial adaptation, $t(38) = 6.2$, $p = 1.4 \times 10^{-7}$ (readaptation for 40-trial washout data: 51.3 ± 4.5%, $t(37) = 5.7$, $p = 9.0$

$\times\ 10^{-7}$ for savings; readaptation for 800-trial washout data: 52.1 ± 4.1%, t(38) = 5.6, $p = 9.9 \times 10^{-7}$ for savings). This indicates substantial, statistically significant savings in temporally-volatile adaptation as illustrated in the bottom row of **Fig 3E and 3F**.

Savings does not arise from the rapid reemergence of temporally-persistent memories

Intriguingly, the clear pattern of savings we found in the learning curves for overall and temporally-volatile adaptation was absent for temporally-persistent adaptation. In only 1 of the 4 conditions in Experiments 1 and 2 (readaptation after a 40-trial washout in Experiment 1) was the unnormalized temporally-persistent adaptation even nominally higher during relearning than initial adaptation, and in that condition the readaptation built upon a substantially higher pretraining level than the corresponding initial training condition (**Fig 3A**). When pretraining levels of persistent adaptation were taken into account by normalizing the learning curves, we found that relearning for temporally-persistent adaptation was nominally slower, rather than faster, than initial learning in all 4 conditions as shown in **Fig 3C and 3D**. Specifically, early (trial 10) savings were, on average −10.0 ± 4.3% of the ideal persistent adaptation, t(38) = −2.3, $p = 0.99$ for savings (40-trial washout data: −7.3 ± 4.4%, t(37) = −1.7, $p = 0.95$; 800-trial washout data: −10.3 ± 4.5%, t(38) = −2.3, $p = 0.99$) as shown in **Fig 3E and 3F**. The temporally-persistent adaptation measured 40 and 70 trials into the training period for the combined 40-trial and 800-trial washout data displays results similar to trial 10 adaptation, with a tendency towards anti-savings (slower readaptation) (t(39) = −3.4, $p = 1.00$ for trial 40 and t(38) = −2.2, $p = 0.98$ for trial 70). The absence of savings in temporally-persistent adaptation stands in stark contrast to the high levels of savings observed in temporally-volatile adaptation, suggesting that overall savings arises from the former, but not the latter. Thus, our result indicates that savings arises from the faster relearning of volatile memories, rather than the re-manifestation of persistent memories.

Temporally-persistent memories display anti-savings

Based on recent work that reported anti-savings for implicit motor adaptation [45], we asked, in a post-hoc analysis, whether temporally-persistent adaptation consistently displayed the slowed relearning that would constitute anti-savings. This analysis revealed that, relearning for persistent adaptation was, in fact, significantly slower than initial learning (t(38) = −2.3, $p = 0.0247$, 2-tailed paired *t* test), based on the trial 10 the data from both washout periods combined. This anti-savings was most clear in the long 800-trial washout data, which allowed us to examine savings without any effects of residual temporally-persistent adaptation (t(38) = −2.3, $p = 0.0276$, 2-tailed

paired *t* test). Savings at trial 10 after the short incomplete washout was also nominally negative but, in this case, not significantly so (t(37) = −1.7, *p* = 0.1073, 2-tailed paired *t* test). Analysis of temporally-persistent adaptation at trials 40 and 70 provides consistent results, with statistically significant anti-savings observed for the combined data from the 40 and 800-trial washout periods (t(39) = −3.4, *p* = 0.0017 at trial 40; t(38) = −2.2, *p* = 0.0359 at trial 70, 2-tailed paired *t* tests) and also for the 800-trial washout data analyzed in isolation (t(39) = −3.3, *p* = 0.0022 at trial 40; t(38) = −3.1, *p* = 0.0035 at trial 70, 2-tailed paired *t* tests). Accordingly, the 40-trial washout data analyzed in isolation showed mixed results at these individual time points (t(39) = −2.1, *p* = 0.0417 at trial 40; t(38) = −0.7, *p* = 0.5053 at trial 70, 2-tailed paired *t* tests). However, when the data combined across all time points are used, we find individually significant anti-savings for both washout periods (t(39) = −5.0, *p* = 0.000012, for the 800-trial data; t(39) = −2.6, *p* = 0.0131 for the 40-trial data). In sum, the negative savings results we observe in the 800-trial and 40-trial washout data are similar, but it appears that the 800-trial result is somewhat clearer, possibly because the 40-trial washout data suffer from incomplete washout of the initial adaptation before relearning. Overall, our data show a conspicuous absence of savings in the relearning of temporally-persistent adaptation in all conditions we examined, instead showing anti-savings despite robust savings in the relearning of temporally-volatile adaptation.

Although a small effect, we found it interesting that anti-savings was somewhat more consistently observed following the 800-trial washout condition than the 40-trial condition, suggesting that the prolonged repeated execution of the same no-rotation trials that constitute the washout period might make anti-savings more consistent. Indeed, the strengthening of an action following repeated execution, often termed use-dependent learning (UDL) [69,70], can manifest in reaching in the form of a directional bias toward its direction [70–72]. Interestingly, the expected bias toward the baseline no-rotation movement direction following washout would oppose the movement direction changes associated with VMR relearning, and thus act in the direction of anti-savings to reduce relearning.

However, it is critical to note that (1) because we are examining savings, the slowed relearning that constitutes anti-savings refers to slower than initial learning; and (2) that the initial learning period in our experiments was also preceded by a prolonged period of the execution of repeated no-rotation trials, which would likewise elicit a UDL effect. The key question would not, therefore, be whether a UDL effect might slow relearning following the 800-trial washout period, but whether such an effect would show a meaningful size increase between the 220-trial duration of the no-rotation baseline period that precedes initial learning and the 800-trial no-rotation washout period that

precedes the 800-trial relearning condition? However, the available literature on how UDL effects increase with the number of repeated trials suggests that this is unlikely. Studies examining UDL effects in reaching movements showed effects after only 1 to 15 trials ([71], Exp 3; [70]), and the one study that looked at the time course for UDL effects beyond 15 trials, found effects that asymptoted between 50 and 150 trials ([71], Fig 4), which is smaller than the duration of the 220-trial baseline that preceded initial learning in our experiment. This suggests that UDL effects, if they indeed affect VMR training in our study, would do so equally for both initial learning, which was preceded by 220 no-rotation trials, and relearning following long-washout, which was preceded by 800 no-rotation trials. Consequently, UDL effects should have little effect on the difference between these learning curves and thus on the savings we measure after 800 washout trials, and are, therefore, unlikely to explain the temporally-persistent anti-savings we observed.
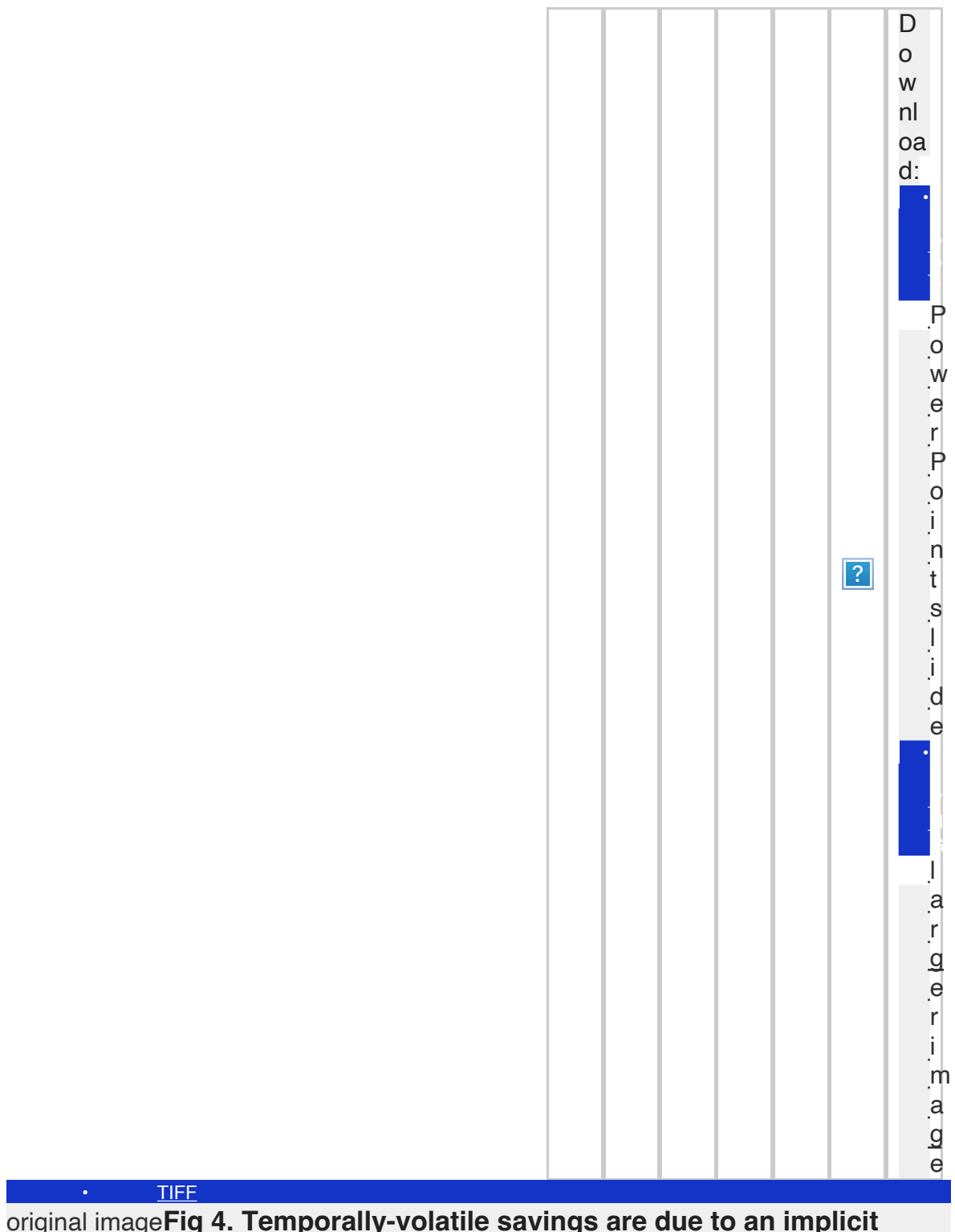
original image **Fig 4. Temporally-volatile savings are due to an implicit adaptation component.**

(a) Learning curves for Experiment 3, showing overall adaptation (solid lines), overall implicit adaptation (square) and persistent adaptation (circles). Gray denotes initial learning, blue denotes relearning. Shading indicates SEM. **(b)** Diagram illustrating measurements of specific components of implicit and explicit learning for Experiment 3, based on a zoomed-in view of the blue

retraining data from panel **a**. Instructions 1 trial before and 1 trial after the 1-min delay (trials 9 and 10) allow the direct measurement of overall-implicit and implicit-persistent components, respectively. This allows measurement of implicit-volatile as the difference between them, and overall-explicit adaptation as the difference between overall adaptation and overall-implicit, where overall adaptation is measured as the average amount of adaptation 2 trials before the first instruction trial and 2 trials after the last persistent-only trial (i.e., trials 7, 8, 12, and 13; blue filled circles). Both instruction trials have no visual feedback to avoid per-trial learning leading to posttrial recovery of adaptation. The following trial (trial 11) thus allows direct measurement of combined implicit and explicit persistent adaptation, therefore, the difference between trial 11 and trial 10 measures explicit-persistent. Trial 11 also reintroduces visual feedback, allowing adaptation to recover by the next trial. **(c)** Savings in overall, temporally-persistent, and temporally-volatile adaptation in Experiment 3, both for combined implicit and explicit adaptation (left column) and broken into implicit and explicit components. In line with Experiments 1 and 2, data show overall and volatile, but not persistent, savings for combined (implicit + explicit) adaptation. Further dissociation into implicit and explicit components reveals this savings is due to implicit, temporally-volatile adaptation. Error bars indicate SEM. *$p < 0.05$; **$p < 0.01$. Underlying data supporting this figure can be found in file Exp_3_data.mat.

https://doi.org/10.1371/journal.pbio.3001799.g004

However, the one paper that looked at the effect of repeat duration [71] did not study UDL training periods as long as 800 trials (540 trials was their maximum) and investigated UDL outside the context of VMR adaptation. We thus performed an additional experiment (Experiment S1) to determine whether the differences between the number of no-rotation reaches before initial learning (220 trials) versus before the long-washout relearning (800 trials) might explain the anti-savings we observe. Experiment S1 examined initial learning after a baseline of 800 rather than 220 trials to match the duration of the action selection history of the 800-trial washout before relearning.

If the reduction in temporally-persistent relearning with respect to initial learning we observed were indeed due to the longer 800 trial movement/action selection history, we would expect the initial temporally-persistent learning in this new dataset to match the slowed temporally-persistent relearning from the 800-trial washout condition rather than the initial temporally-persistent learning that we previously observed after 220 baseline trials. However, the results from Experiment S1 instead show that the initial temporally-persistent learning following 800 baseline trials in this new dataset was a closer match to the initial temporally-persistent learning that was previously observed after 220 baseline trials, as this dataset did not show the

slower initial temporally-persistent learning that would be predicted by increased UDL following 800 rather than 220 trials (trial 10 learning: 47.1 ± 9.6% for initial learning in the new data versus 40.1% ± 4.2% for initial learning in the previous data, t(49) = –0.8, p = 0.45, see **S1 Fig**). Instead, as shown in **S1 Fig**, the data are in line with the Verstynen and Sabes data [71] whereby UDL effects asymptote before 220 trials. In line with this prediction, the temporally-persistent relearning following 800-trial washout trials observed in the previous data was also significantly slowed compared to this new 800-trial baseline data with significant anti-savings observed when the data from trials 10, 40, and 70 were averaged together (t(50) = –3.2, p = 0.0014) and also when the data from these time points were analyzed separately (t(50) = –2.2, p = 0.017 at trial 10; t(50) = –2.5, p = 0.0086 at trial 40; t(50) = –2.2, p = 0.0162 at trial 70). These findings suggest the UDL effects cannot explain the 800-trial temporally-persistent anti-savings we observe.

As a side note, if UDL effects were, on the other hand, somewhat lower after 40 trials (the short washout period duration) compared to 220 trials (the baseline period duration) [71], it would lead to a reduction in the amount of UDL-induced slowing for relearning after 40 washout trials compared to the initial learning. This would suggest that, if anything, we underestimated temporally-persistent anti-savings in the 40-trial washout data rather than overestimating it in the 800-trial washout data, perhaps contributing to the less consistent results when this condition was considered in isolation.

**Temporally-volatile savings arise from implicit adaptation**

Previous research associated savings in visuomotor adaptation with the rapid recall of explicit strategies, rather than faster implicit adaptation [18,43,44]. This led us to investigate the contributions of implicit and explicit processes in the temporally-volatile savings we observed in our paradigm. We thus ran Experiment 3 (N = 40), which consisted of two 80-trial learning episodes separated by 800 washout trials. We dissected savings into implicit and explicit components using special instruction trials that prompted participants to disengage any explicit strategy by aiming their hand directly to the target [44,73–78]. These instructions were presented immediately before and after the first (trial 10) 60-s time delay following the onset of the VMR in both initial learning and relearning and allowed us to dissect adaptation into 4 subcomponents: implicit-persistent, implicit-volatile, explicit-persistent, and explicit-volatile (**Fig 4B**, see Materials and methods for details).

In line with our findings in Experiments 1 and 2, we found savings for overall and volatile adaptation (14.3 ± 3.6%, t(39) = 4.0, p = 0.00014 and 11.2 ± 4.8%, t(37) = 2.4, p = 0.0119, correspondingly) but not persistent adaptation (4.0 ± 4.9%, t(38) = 0.8, p = 0.21). Dissection of savings into explicit and

implicit components revealed savings for both overall implicit and implicit-volatile adaptation (14.1 ± 5.7%, t(38) = 2.5, p = 0.0088 and 13.3 ± 5.5%, t(37) = 2.4, p = 0.0104, correspondingly) but not explicit-volatile adaptation (−2.1 ± 6.2%, t(37) = −0.3, p = 0.63) or any of the persistent subcomponents (implicit-persistent: 3.8 ± 4.7%, t(38) = 0.8, p = 0.21; explicit-persistent: 0.2 ± 4.4%, t(38) = 0.1, p = 0.48). This finding suggests that overall savings were driven by the implicit and temporally-volatile component of adaptation, in turn suggesting that the temporally-volatile savings we observed in Experiments 1 and 2 predominantly reflect an implicit process rather than an explicit strategy. That the volatile component observed in Experiments 1 and 2 is primarily implicit is not surprising: First, it is unclear why an explicit strategy could be temporally-volatile to the point of being largely or completely forgotten after a short 1-min delay. In fact, our recent work indicates that explicit adaptation displays essentially no temporal volatility, with over 95% stability across 1-min delays [79]. Second, our paradigm elicited scant explicit adaptation (likely due to elements of our experiment design aimed at inducing implicit learning such as the use of point-to-point (rather than shooting) movements, the lack of aiming instructions, the lack of markers that could aid off-target aiming, and the presence of low-latency online feedback [76,80–83]) and without substantial explicit adaptation we lacked power for measuring explicit savings.

Dissecting long-term memory in visuomotor adaptation

We next investigated whether the ability to dissect motor learning into temporally-persistent and temporally-volatile components could shed light on the mechanisms for the formation of long-term memories. To accomplish this, we examined the relationship between the levels of temporally-persistent and temporally-volatile learning observed after initial training and the amount of retention observed 24 h later (Experiment 4). After a baseline period, we trained 25 participants on a 30° VMR for 120 trials. After this initial training, they were tested for temporally-persistent adaptation as present after a rest break (average break duration: 125 ± 8 s, which would let >99% of temporally-volatile adaptation decay based on a time constant of approximately 20 s). The above measurements were then repeated, with participants retrained for 60 trials, and retested for temporally-persistent adaptation (the average of these 2 measurement sessions was used to quantify temporally-persistent adaptation for each individual). Participants then returned the following day to be tested for retention (**Fig 5A**, see Materials and methods).
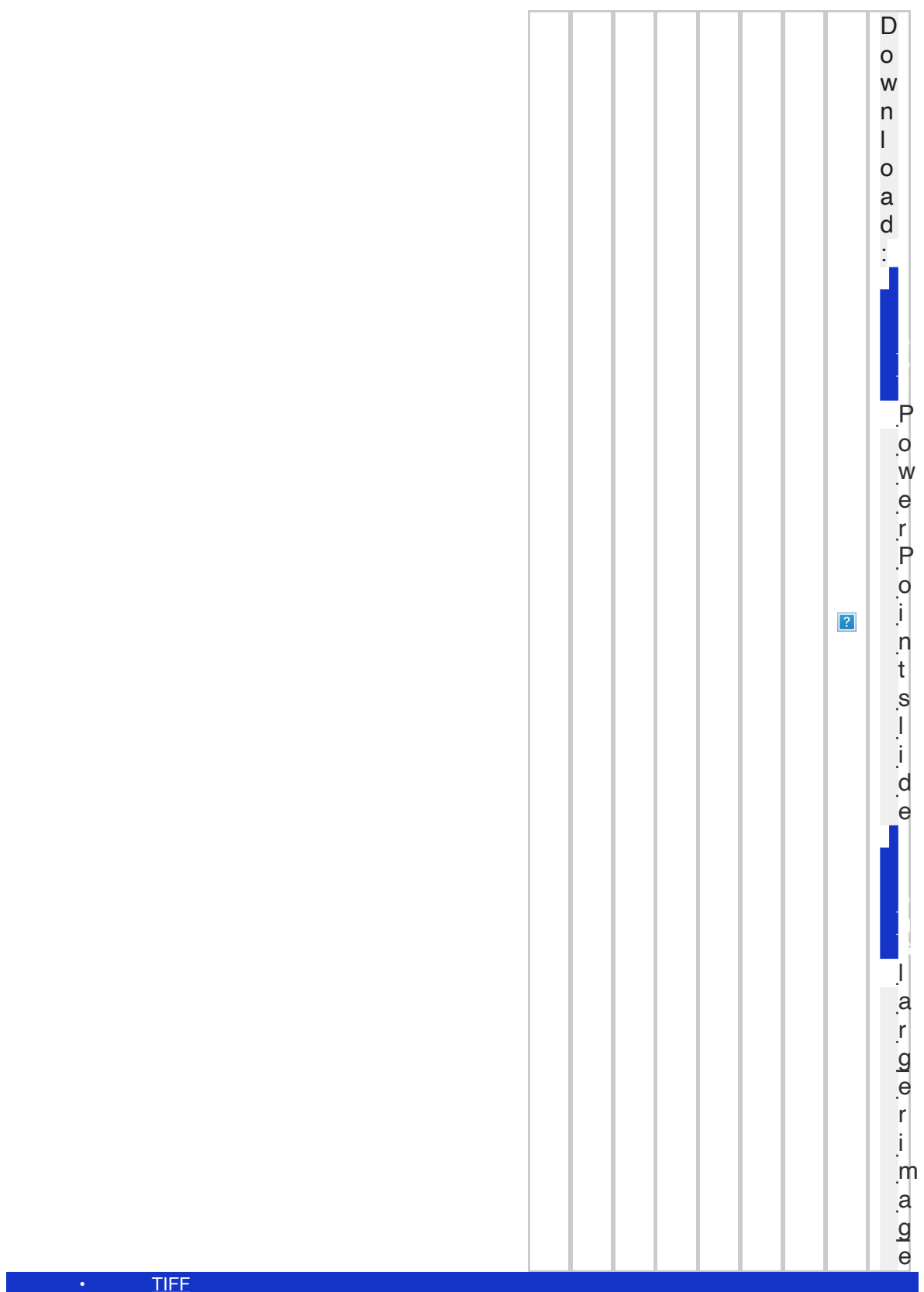
original image**Fig 5. Measuring temporally-volatile and temporally-persistent components of adaptation and subsequent long-term retention.**

(a) Experiment schedule and raw data for Experiment 4. After a baseline period, subjects were trained with a 30° VMR for 120 trials, and were then tested, after a break, for temporally-persistent adaptation, retrained for 60 trials, and then retested after another break. Subjects returned the following day when they were tested for 24-h retention (orange circle). Note that 24-h retention is lower than overall or temporally-persistent adaptation but higher than zero. TP: temporally-persistent; TV: temporally-volatile. Yellow background indicates trials used to measure overall adaptation. Error bars and shading indicate SEM. **(b)** Comparison of overall, temporally-persistent, and temporally-volatile adaptation from Experiments 1, 2, and 4 with the 24-h retention from Experiment 4. Experiments 1, 2, and 4 display similar levels of persistent and volatile adaptation. Underlying data supporting this figure can be found in file Exp_4_data.mat.

https://doi.org/10.1371/journal.pbio.3001799.g005

We found that, the overall adaptation measured late in training (the last 20 trials) in Experiment 4 was similar to that observed in Experiments 1 and 2 (27.4 ± 0.3° for Experiment 4 versus 26.4 ± 0.6° and 27.7 ± 0.5° for Experiments 1 and 2, see **Fig 5B**). Similarly, the persistent component of adaptation was also similar across the 3 experiments (16.7 ± 1.0° for Experiment 4 versus 18.1 ± 0.7° and 20.4 ± 1.0° for Experiments 1 and 2, see **Fig 5B**), suggesting that the somewhat longer training duration in Experiment 4 had little effect on either overall or temporally-persistent adaptation. When examining long-term memory, retained 24 h after training, we found that participants retained 8.9 ± 1.1° of the trained 30° rotation (orange bar in **Fig 5B**). This corresponded to 32.4 ± 4.2% of the overall learning and 52.7 ± 5.2% of the temporally-persistent learning from day 1.

Dissociable effects of temporally-volatile and temporally-persistent adaptation on the formation of long-term memory

To examine whether the dissection of day 1 learning into temporally-persistent and temporally-volatile components could shed light on the mechanism for long-term motor memory formation, we compared the levels of temporally-volatile, temporally-persistent, and overall learning to the amount of 24-h retention for each individual participant. Looking for positive contributions of each component to 24-h retention (using linear regression with regression coefficients restricted to be positive), we found no significant relationship between overall learning on day 1 and 24-h retention on day 2 ($r = +0.14$, $F(23,1) = 0.4$, $p = 0.51$). However, we found a highly significant positive relationship between persistent learning on day 1 and 24-h retention (slope = 0.80, $r = +0.71$, $F(23,1) = 22.9$, $p = 0.00008$). In contrast, we found no positive relationship between volatile learning on day 1 and 24-h retention; in fact, the best fit slope was zero ($r = 0.0$, $F(23,1) = 0$, $p = 1$), as the best fit slope without

restricting regression coefficients to positive values would have been negative. This indicates that temporally-volatile learning does not lead to 24-h retention, consistent with the fact that volatile learning, by definition, will decay over the course of 1 min. We thus find that, whereas neither overall adaptation nor the temporally-volatile component can predict it, the temporally-persistent component of adaptation, measured only 1 min after training, is able to accurately predict retention 24 h after training.

We next performed a stepwise bivariate regression analysis of how 24-h retention was associated with temporally-volatile and temporally-persistent learning from day 1, as illustrated in **Fig 6A and 6B**. This analysis was particularly important here because temporally-volatile and temporally-persistent learning were not independent across individuals but instead displayed a strong negative relationship such that participants with higher day 1 temporally-volatile learning displayed smaller day 1 temporally-persistent learning and vice versa. This bivariate regression revealed that adding temporally-volatile learning as a second regressor after temporally-persistent learning resulted in no significant improvement in the ability to explain 24-h retention (R^2 increased from 49.8% to 51.7% corresponding to a partial R^2 of only 3.8%, $F(22,1) = 0.9$, $p = 0.36$). In contrast, adding temporally-persistent learning as a second regressor after temporally-volatile learning resulted in a large improvement in the ability to explain 24-h retention (R^2 increased from 0.0% to 51.7%, corresponding to a partial R^2 of 51.7%, $F(22,1) = 23.6$, $p = 0.00007$). The results of this analysis are shown in **Fig 6A and 6B** where we illustrate the partial R^2 analysis by comparing each component of day 1 learning with the portion of 24-h retention not explained by the other (see Materials and methods for details). When we repeated this analysis using estimates of temporally-persistent and temporally-volatile adaptation based on either the first or the second measurement session alone rather than the averaged data, we found similar results (Session 1 only: partial R^2 of 43.1%, $p = 0.0005$ for temporally-persistent adaptation versus partial R^2 of 1.0%, $p = 0.64$ for temporally-volatile adaptation; Session 2 only: partial R^2 of 53.0%, $p = 0.00005$ versus partial R^2 of 8.6%, $p = 0.16$, correspondingly). This indicates that the measurements of temporally-persistent learning from both day 1 sessions independently predict subsequent 24-h retention on day 2, albeit with a nominally stronger association for the second session which was adjacent to the 24-h retention measurement.
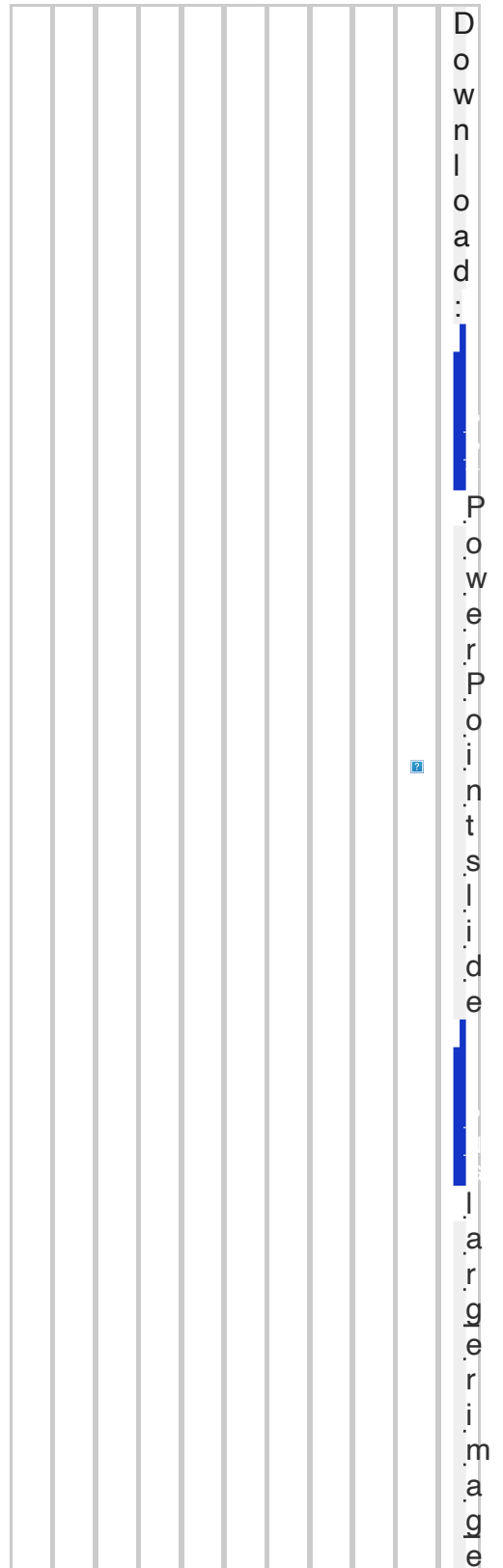
Fig 6. A double dissociation between savings and long-term memory, uncovered by dissection of learning into temporally-persistent and temporally-volatile components.

(a) Illustration of the partial regression between temporally-persistent adaptation on day 1 (shown on x-axis) and 24-h retention across individuals ($N = 25$). The y-axis represents residuals of the univariate regression of 24-h retention upon temporally-volatile adaptation. The positive relationship indicates that higher temporally-persistent adaptation is associated with higher long-term memory. The solid line indicates linear fit. (b) Same as **a** but for temporally-volatile adaptation, showing no significant relationship (slopes restricted to positive values). (c) Illustration of the partial regression between temporally-persistent adaptation during relearning (shown in x-axis) and savings for early (trials 2–6) training combined across Experiments 1 and 2. The y-axis represents residuals of the univariate regression of savings upon temporally-volatile adaptation. There is no significant relationship. (d) Same as **c** but for temporally-volatile adaptation, showing a significant positive relationship and thus indicating that temporally-volatile adaptation explains overall savings. Note that mean y-axis values in (a)–(d) are zero, since they represent residuals of linear regression; these values do not reflect the actual amounts of long-term retention or savings which are, on average, significantly above zero. (e) Comparison of average savings for temporally-persistent (TP) and temporally-volatile (TV) adaptation, combined across Experiments 1 and 2. Error bars indicate SEM. ***$p < 0.001$. Underlying data supporting this figure can be found in files Exp_1_2_data.mat and Exp_4_data.mat.

In summary, we find in Experiment 4, that increased temporally-persistent adaptation is associated with stronger long-term memory, whereas increased temporally-volatile adaptation does not. This sharply contrasts with Experiments 1 and 2 where we found that temporally-volatile adaptation was associated with savings whereas temporally-persistent adaptation was not. Taken together, these results demonstrate a striking double dissociation between the contributions of temporally-persistent and temporally-volatile learning to long-term memory and savings. To even more directly compare the 2 contributions to this double dissociation, we returned to the savings data from Experiments 1 and 2 and performed a bivariate analysis of the inter-individual differences in overall savings based on the levels of temporally-volatile and temporally-persistent learning during retraining (see Materials and methods for details). This bivariate regression analysis is analogous to that performed on the 24-h retention data above and is illustrated in **Fig 6C and 6D**, in parallel format to **Fig 6A and 6B**. Adding temporally-volatile learning as a second regressor after temporally-persistent learning resulted in a significant improvement in the ability to explain savings (R^2 increased from 0.0% to 13.8% corresponding to a partial R^2 of 13.8%, $F(37,1) = 5.9$, $p = 0.0198$). We note here that this p-value, corresponds to the variance reduction associated with adding temporally-volatile learning to the regression and is equivalent to a two-tailed test for its regression slope being non-zero. Thus, if a one-tailed test

would have instead been used, in line with the idea of testing for a significantly positive relationship between savings and initial temporally-volatile learning, its *p*-value would have been <0.01. In contrast, adding temporally-persistent learning as a second regressor after temporally-volatile learning resulted in no significant improvement in the ability to explain savings (R^2 increased from 13.7% to 13.8%, corresponding to a partial R^2 of 0.2%, F(37,1) = 0.1, *p* = 0.81). The findings from this regression analysis show that interindividual differences in the temporally-volatile but not the temporally-persistent component of initial learning are associated with individual differences in the amount of savings during relearning. This adds to the evidence illustrated in **Fig 3**, summarized in **Fig 6E,** that temporally-volatile learning displays savings whereas temporally-persistent learning does not.

Discussion

Here, we compared the mechanisms responsible for savings and long-term memory in human motor learning, finding that temporally-volatile adaptation leads to savings and that temporally-persistent adaptation leads to long-term memory. When we dissected adaptive responses into temporally-persistent and temporally-volatile components using 60-s delays (**Fig 1**), we found that temporally-persistent memories washed out 4 to 20× more slowly than temporally-volatile memories (**Fig 2**), leaving a considerable temporally-persistent residual even after 100 washout trials (see **Fig 2**), when overall adaptation had long since washed out to zero. This suggests that the short washout periods of 11 to 100 trials per trained movement direction used in a number of previous studies to wash out overall adaptation [13,15,16,18,24, 64] would likely have failed to wash out this temporally-persistent adaptation.

We, therefore, controlled for the effect of residual temporally-persistent adaptation, either experimentally with an extended 800-trial washout period which could eliminate it, or analytically with appropriate baseline subtraction and normalization, allowing us to accurately assess savings. With this control, we consistently found significantly greater savings for temporally-volatile than temporally-persistent adaptation (**Figs 3 and 6**). In fact, whereas temporally-volatile adaptation showed savings by displaying relearning that was consistently faster than initial learning, temporally-persistent adaptation remarkably showed anti-savings—displaying relearning that was significantly slower than initial learning both in the overall data and specifically in the 800-trial washout condition in which complete washout of both persistent and volatile learning occurred, allowing savings to be most cleanly measured. Temporally-persistent learning was also nominally slower, though not significantly so, in the 40-trial washout data. Remarkably, we found that

savings in temporally-volatile adaptation was sufficiently large to overcome the anti-savings in persistent adaptation, and still confer robust savings on overall adaptation. Moreover, we found that the temporally-volatile savings we observed were due to implicit rather that explicit learning (**Fig 4**). Our data thus suggest that savings in overall adaptation is derived from implicit, temporally-volatile adaptation, representing an increased propensity to more rapidly form a temporally-volatile memory, rather than the reemergence of a temporally-persistent memory.

When we dissected adaptation into volatile and persistent components to examine the mechanisms for long-term memory, we found a strong positive relationship between 24-h retention and temporally-persistent but not temporally-volatile adaptation (**Figs 5 and 6**). Together, our findings for savings and 24-h retention delineate a powerful double dissociation whereby temporally-persistent learning leads to long-term memory but not savings and temporally-volatile learning leads to savings but not long-term memory.

The juxtaposition between savings in temporally-volatile and temporally-persistent adaptation explains differences in savings across previous studies

Our findings provide a resolution to the apparent discrepancy between previous studies which isolated implicit savings in adaptation, yet found either anti-savings [45] or savings [20]. The paradigm used in the former study likely promoted temporally-persistent implicit adaptation, which here we find to display anti-savings, because compared to the current findings which employed only 1 movement direction, the multiplicity of movement directions used would dramatically increase the temporal spacing between same-direction movements, limiting temporally-volatile adaptation—though the extent of this effect is unfortunately difficult to definitively evaluate because the inter-trial time intervals were not explicitly reported in the study. In contrast, the quickly paced paradigm used in the latter study would permit greater accumulation of temporally-volatile adaptation, which here we find to display enough savings to overcome anti-savings in temporally-persistent adaptation. Other factors may also have affected the balance between temporally-volatile and temporally-persistent components to drive the savings versus anti-savings observed in these 2 studies.

This effect whereby temporally-volatile savings would be reduced when a larger number of target directions are present in the experiment design, because same-direction inter-trial time intervals would increase and force temporally-volatile adaptation to decay to a greater extent, would also predict reduced savings even in studies that did not isolate implicit adaptation. This prediction is indeed borne out in previous work, with studies using 4- to 8-target paradigms finding either less pronounced savings [13,19] or no savings

at all [24], whereas studies using 1-target or 2-target paradigms [15,17,18] demonstrating considerable savings. Given the approximately 20 s time constant observed for the decay of temporally-volatile adaptation during VMR learning [56–58], the fairly rapid 3-s inter-trial interval in our study would allow near-complete (86%) carryover of temporally-volatile adaptation from one trial to the next, whereas the 8-fold increase in inter-trial intervals expected for an 8-target study with similar trial pacing would allow only 30% carryover. This would dramatically reduce the effect of temporally-volatile savings.

Incomplete washout can contaminate the assessment of savings

Our experiments revealed that temporally-persistent adaptation requires a surprisingly long period to wash out—well above 100 trials and much longer than overall adaptation that combines volatile and persistent components and is effectively washed out in just 20 to 40 trials (**Fig 2**). This occurs because a negative temporally-volatile adaptation acts to mask the enduring temporally-persistent component after 20 to 40 washout trials. Therefore, if the washout of temporally-persistent adaptation is not specifically measured, it is easy to get the false impression that a short block of 20 to 40 trials is sufficient to washout all adaptation so that true savings—which refers to relearning after complete washout of all adaptation—can be cleanly measured. In fact, most savings studies we examined employed washout periods well below what would be necessary for complete washout of temporally-persistent adaptation [13,15,16,18,24,64]—a notable exception being a recent [45] study which employed a long washout period with several breaks that probed for residual temporally-persistent adaptation. Incomplete washout of temporally-persistent adaptation would lead to an apparent savings due to the unmasking of this persistent memory during relearning as predicted by a multi-rate learning model [10], contaminating the assessment of true savings. Even Zarahn and colleagues [15], where the primary claim was that savings was present after the complete washout of all components of adaptation so that the savings they found could not be explained by apparent savings due to the unmasking of a slowly decaying learning process, used only 40 trials, meaning that the savings they observed almost certainly included apparent savings from incomplete washout of temporally-persistent adaptation. In fact, inspection of the Zarahn and colleagues data reveals that savings at trial 10 (the time point at which we found maximal savings) was about 20% of the perturbation size. However, we find that the short 40-trial washout that they used resulted in a residual level for temporally-persistent adaptation that was about 15% of the perturbation size, suggesting that the most of the savings they observed was apparent savings. The current results from the 800-trial washout data where temporally-persistent adaptation was entirely eliminated, do, however, show clear evidence for the true savings that Zarahn and colleagues hypothesized. It is unfortunately impossible to know precisely how much of the savings that

study reported was due to incomplete washout because of paradigmatic differences, in particular their use of a 45-degree rather than a 30-degree VMR perturbation, that may have increased or decreased the degree to which their washout was incomplete or the amount of true savings present.

Implicit savings does not result from the recall of a consolidated motor memory

Our findings dissociating savings from long-term memory, and demonstrating savings to be driven by temporally-volatile, rather than temporally-persistent memories, upend the widespread view that the faster relearning that characterizes savings results from the recall of a previously consolidated, stable motor memory [8,9,14,16,17,19,43,84,85]. Whereas savings has been taken as a litmus test for the consolidation of motor memory 1, 2, or even 7 days after training [13,24], here we find that savings is primarily driven by the temporally-volatile component of adaptation, which decays over the course of 1 min or less. These temporally-volatile savings could not arise from a consolidated memory of previous adaptation, as any memory solid enough to survive the hour-long, 800-trial washout period in our experiments, would certainly be stable against the passage of time during the minute-long delays in our experiment used to define temporal lability. Our findings thus indicate that savings is not driven by a consolidated memory of the adapted motor output, in line with the double dissociation between savings and long-term memory that we demonstrate.

In contrast, the savings we observe is driven by an increased learning rate for adaptive but temporally-volatile changes in motor output from one trial to the next when experiencing the same perturbation again after washout. While the ability to establish this learning rate increase could be considered a type of consolidated memory as it is long-lasting and must rely on a plastic change driven by experience, this ability is very different from the usual conception of a consolidated motor memory in which the trained actions themselves are remembered. Thus, our findings are consistent with a model in which implicit savings is based on an ability to improve the adaptation of actions, rather than a memory for the actions themselves. It is important to point out, however, that our findings do not show that consolidated memories are not formed or eliminate the possibility that they make some contribution to savings. For example, it is possible that explicit savings is driven by consolidated explicit memory. Even in our implicit adaptation data for which temporally-persistent relearning displays anti-savings, we cannot rule out the possibility that a consolidated temporally-persistent memory subcomponent exists and contributes to savings but that that the effect is overwhelmed by a larger anti-savings effect from a different subcomponent. Instead, our findings

demonstrate that consolidated memories are not required for savings.

Mechanisms for learning rate modulation in temporally-volatile adaptation

What are the mechanisms behind learning rate increases in temporally-volatile adaptation? Recent work suggests that such a learning rate increase can be driven by learning environments with increased statistical consistency, defined as a positive correlation between successive perturbations (i.e., lag-1 autocorrelation) in environmental dynamics or errors from one trial to the next [52–54]. This consistency-driven effect is further enhanced when the environment repeats the same perturbation [54], with highly consistent, repetitious switching environments increasing learning rates up to 3× from baseline and highly inconsistent ones decreasing then up to 5×. Critically, the initial training periods in savings paradigms, including the current one, are characterized by both consistency and repetition, as they usually consist of a large number of trials with the same perturbation, leading to an increase in the learning rate during retraining.

Savings and explicit adaptation

Experiment 3 revealed that the temporally-volatile savings we observe arises from implicit adaptation. This adds to recent evidence [19,20] against the idea that savings is exclusively driven by explicit adaptation [18,43–45]. Although savings can clearly occur because of explicit strategies, we did not observe this in our experiment. While this may, in part, reflect the considerable variability in the balance between implicit and explicit adaptation across individuals [74,86], the lack of explicit savings in our study was likely due to an experimental design that promoted implicit adaptation and minimized explicit strategy, and thus provided little power to detect explicit savings. The design elements included the lack of aiming instructions, the absence of workspace markers positioned to aid re-aiming, the use of point-to-point rather than shooting movements, and the engineering of low (approximately 25 ms) visual feedback latency for onscreen cursor motion, all of which may promote implicit learning [76,80–83]. In contrast, studies which reversed most of these design elements elicited primarily explicit adaptation and found clear explicit savings [44].

Taken together, the evidence now suggests that both implicit and explicit adaptation can show savings, albeit via different mechanisms. The current study shows that implicit savings is specifically driven by the faster relearning of a temporally-volatile memory, whereas previous work provides evidence that explicit savings is driven by temporally-persistent memory, as explicit savings is clearly observable in multi-target paradigms that would minimize temporally-volatile memory [44,45]. As a consequence, for experimental

paradigms that primarily elicit explicit learning, we would expect savings to primarily be driven by this temporally-persistent adaptation, and for experimental paradigms that elicit a balance of implicit and explicit learning, we would expect the dichotomy between temporally-persistent and temporally-volatile contributions to savings to be blurred. However, because we did not examine these cases, we cannot know whether our expectations will be borne out or whether complex interactions between implicit and explicit learning might lead to different results that cannot be predicted from the current findings.

Parallels between temporally-volatile/temporally-persistent learning and the fast/slow learning processes of motor adaptation

Another line of work has dissected motor adaptation, not experimentally, but instead on the basis of a computational model with 2 distinct adaptive processes: a fast adaptive process that learns rapidly and displays weak retention, and a slow adaptive process that learns slowly and displays strong retention [10]. By manipulating the training duration in order to elicit different amounts of the fast and slow learning, a subsequent study found that 24-h retention was specifically predicted by the amount of slow learning, rather by the amount of fast learning or overall adaptation [60]. Interestingly, this model-based dissection mirrors our temporal-stability–based dissection as the slow process, like temporally-persistent adaptation, leads to 24-h retention, whereas the fast process, like temporally-volatile adaptation, does not. In fact, there is a remarkable correspondence between the Joiner and colleagues study, which found that 49 ± 6% (95% confidence) of slow learning on day 1 is retained after 24 h, and Experiment 4 in the current study, which found that 46 ± 9% of persistent learning on day 1 is retained after 24 h. Moreover, the trial-to-trial learning characteristics of the fast and slow processes mirror the ones for volatile and persistent adaptation, respectively. In particular, slow adaptation displays slower learning and better retention than fast adaptation, just as temporally-persistent adaptation displays slower learning and better retention than temporally-volatile adaptation (see **Figs 3C, 3D, and 2A**, respectively). These parallels argue, although speculatively so, that the temporally-volatile and temporally-persistent learning from our dissection of adaptation corresponds to the implicit components of fast and slow processes from the two-state model. This possibility challenges 2 prominent ideas from the recent literature. First, the possible correspondence between the fast process and implicit temporally-volatile learning challenges the idea that fast process learning is synonymous with explicit adaptation [18,43,87]. Second, the possibility of a measureable instantiation of fast and slow process learning from the two-state model challenges the assertion that models of context-based learning and the switching between should supplant models of adaptive

processes with different learning rates [41].

The coexistence of temporally-volatile and temporally-persistent memories provides a mechanism for contextual interference

The possible mapping of temporally-volatile adaptation onto a fast learning/ low retention process and of temporally-persistent adaptation onto a slow learning/high retention process provides an intriguing potential explanation for previous work on contextual interference in both motor and cognitive tasks. Contextual interference refers to phenomenon that memories formed in high-interference environments, where the task being performed is randomly switched from one trial to the next, are learned more slowly but show higher retention than memories formed in low-interference environments, where a single task is serially practiced [88–93]. If, as in the VMR adaptation task studied here, both temporally-volatile and temporally-persistent memories contribute to learning in tasks where contextual interference has been observed, then contextual interference effects can be predicted based solely on the temporal spacing inherent in the paradigms that elicit it. The idea here is that the high-interference condition in which tasks are randomly intermingled from one trial to the next would necessarily increase the temporal spacing between the trials within each task compared to the low-interference condition in which tasks are serially practiced. This increase in temporal spacing would allow temporally-volatile memories to decay, at least partially, and thus reduce the amount of temporally-volatile learning, which would slow the overall learning and promote increased temporally-persistent learning. Moreover, the resulting reduction in temporally-volatile learning and increase in temporally-persistent learning would both act to increase the proportion of learning that is temporally-persistent in the high-interference condition, which would, in turn, increase long-term retention according to the current findings. The slowed overall learning and increased retention predicted here for the high-interference condition are, in fact, the defining features of contextual interference. Correspondingly, the converse, faster overall learning but reduced retention, would be predicted for the low-interference condition where serially practiced tasks with reduced temporal spacing would allow temporally-volatile memories to rapidly build during training to improve performance, but would decay before a retention or transfer test, resulting in poor retention. Thus, the coexistence of temporally-volatile and temporally-persistent memories provides an explanation for contextual interference that does not require any interference itself. Further work will be required to determine the fraction of observed contextual interference effects that stem from this mechanism.

Materials and methods

Ethics statement

This study was approved by the Harvard University Committee on the Use of Human Subjects (CUHS). Participants were naïve with respect to the purpose of the experiments and provided written informed consent in accordance with CUHS policies.

Participants

A total of 118 subjects (48 men, age 22.6 ± 4.7, 13 left-handed) participated in the present study (20 each in Experiments 1 and 2, 12 in Experiment S1, 41 in Experiment 3, and 25 in Experiment 4).

Apparatus

We used the same experimental setup as the one used in recent work [62,63]. Subjects sat in front of an apparatus consisting of a 200 Hz digitizing tablet (Wacom Intuos 3 12" × 19", resolution of position data: 0.005 mm; accuracy: 0.25 mm) positioned below a 23" 120 Hz LCD monitor. During the experiment, subjects moved a custom-made handle, which contained a stylus, on top of the tablet allowing us to record hand position. Vision of the hand was occluded by the monitor and subjects instead observed their movement on the screen through a white cursor representing hand position.

Experiment protocol

Using their dominant hand, subjects made point-to-point arm reaching movements between a starting position and targets 9 cm away. At the end of each movement, they were rewarded with a bell sound if they had managed to reach and stop at the target within 250 ms. Training was isolated to the outward movements, as visual feedback was unavailable during the return movements after the first 110 trials of the baseline block. Subjects took rest breaks roughly every 200 trials (about 7 to 10 min, see **Fig 1**).

Experiments 1 and 2 consisted of reaches towards a 90° target direction (in the midline, directly away from the body). After the 220-trial baseline block with no visual rotation, subjects in Experiment 1 ($N = 20$) entered the main part of the session which contained three 80-trial training periods. During training, a 30° VMR was imposed about the starting position. The sign of this VMR was the same for all training periods for each subject, with half the subjects training with a clockwise VMR and the other half training with a counter-clockwise VMR. The first and second training periods were separated by a 40-trial washout period, whereas the second and third training periods were separated by an 800-trial washout period. The training schedule in

Experiment 2 (*N* = 20) was the same apart from that the 800-trial washout period came first (between the first and second training periods, see **Fig 1**).

Experiment S1 (*N* = 12) consisted of a longer, 800-trial baseline block with no visual rotation followed by a single 80-trial VMR training period like the ones used in Experiments 1 and 2.

Experiment 3 (*N* = 41) was similar to Experiment 2 in that it contained two 80-trial training periods separated by a 800-trial washout period (but not a third training period). It was designed to examine whether the temporally-volatile savings like the ones observed in Experiments 1 and 2 were due to an implicit or explicit process.

To dissect savings into implicit and explicit components, we used special instruction trials that prompted participants to disengage any explicit strategy by aiming their hand directly to the target. This method, also referred to as exclusion (since participants are to exclude strategies from their reach) [94], has been, in various forms, widely used to dissect implicit and explicit visuomotor adaptation [44,73–78]. Specifically, instructions were given to either move to the center of the target or to its near/far end (both of which would not alter the reaching angle) and were presented immediately before and after the first (trial 10) 60-s time delay within both VMR training episodes (initial learning and relearning).

This enabled us to directly assess overall implicit adaptation (the amount of adaptation on the first instruction trial) and implicit-persistent adaptation (the amount of adaptation in the second instruction trial, which followed a 60-s delay), and, by comparing these 2, this enabled us to assess implicit-volatile adaptation. Both instruction trials had no visual feedback to avoid per-trial learning that would lead to posttrial recovery of adaptation. Moreover, by comparing adaptation in the second instruction trial to the no-instruction trial following it, we assessed explicit-persistent adaptation, and, by estimating overall adaptation as the average adaptation 2 trials before and after all these delay/instruction trials, we obtained estimates of overall explicit, volatile, and persistent adaptation (**Fig 4B**).

To minimize delays in reaction time, which would increase the inter-trial time interval and lead to further reduction in temporally-volatile adaptation, participants were presented with an "upcoming instruction" sound during the trial preceding the instruction. To familiarize participants with instruction trials (and the preceding "upcoming instruction" sound) ahead of VMR training, we presented a series of similar instruction trials during familiarization. Familiarization contained 4 different possible instructions: move your hand to

the near, far, left, or right end of the (circular) target. There were clear biases towards the instructed endpoints showing adherence to the instructions.

The aim of Experiment 4 ($N = 25$) was to examine the formation of long-term memories of VMR adaptation. The experiment began with a baseline period with no VMR that consisted of 456 trials, spread evenly across 19 target directions. After this baseline, subjects were trained on a 30° VMR for 120 reaches to a target placed at 90° (in the midline, directly away from the body, the same target used in Experiments 1 to 3). The direction of the 30° visual rotation was approximately balanced, with 13 subjects trained with a counter-clockwise VMR and 12 subjects with a clockwise VMR. This was followed by a testing block with 3 reaches towards each of the 19 targets, including the 1 target direction used throughout Experiments 1 to 3 (the other 18 directions were sampled to assess generalization of VMR learning as part of a separate study; here, we focus on learning and retention along the trained direction). During this block, visual feedback was withheld so that repeated measurements could be made without these measurements being contaminated by additional training that could be elicited by visual feedback. We used the movements towards the training direction to measure temporally-persistent adaptation. After this testing block, subjects were retrained on the 30° VMR for an additional 60 trials and after that were tested again without visual feedback to measure temporally-persistent adaptation as described above. Participants returned the following day to be tested for 24-h retention without visual feedback.

Sample size determination

While sample sizes for experiment groups in analogous studies typically range between 8 and 12, here, we used somewhat larger sample sizes ($N = 20, 20, 41$, and 25 for Experiments 1, 2, 3, and 4, respectively; Experiment S1 had 12 participants, mirroring sample sizes in analogous studies). For Experiments 1 and 2, we examined a larger number of participants so that we could rigorously assess not only whether savings is present or not for temporally-persistent and temporally-volatile adaptation, but also the time course of savings for these 2 adaptation components at multiple points during training, as well as whether there are any subtle differences in savings or the extent of washout following the 40-trial versus the 800-trial washout periods. The larger sample sizes in Experiments 1 and 2 also enabled more precise comparisons between the time course of washout for temporally-persistent and temporally-volatile adaptation, as the time constant estimates for these washout curves can be especially susceptible to noise in the data. In Experiment 3, we doubled the sample size relative to Experiment 2, given that Experiment 3 involved dissection of adaptation into 4 (explicit-persistent, explicit-volatile, implicit-persistent, and implicit-volatile), rather than 2 components. In

Experiment 4, we examined $N = 25$ participants as we wanted to be able to look at not just the group-average amount of 24-h retention, but also examine how inter-individual differences in 24-h retention on day 2 related to inter-individual differences in temporally-persistent and temporally-volatile adaptation on day 1 (**Fig 4B and 4C**).

Data analysis

Statistical comparisons.

We performed single-sided paired *t* tests across subjects to assess the presence of (positive) savings in adaptation and its subcomponents. For Experiment S1, which was designed to investigate potential mechanisms for anti-savings in temporally-persistent adaptation, we similarly used single-sided unpaired *t* tests to compare Experiment S1 initial learning data to Experiment 1/2 800-trial washout relearning data. For all other statistical comparisons two-sided paired *t* tests across subjects were implemented, with the exception of the comparisons involving the estimation of washout time constants in **Fig 2A** and the estimation of confidence intervals associated with the % contribution of temporally-persistent or temporally-volatile savings to overall savings: In these cases, we used a bootstrapping procedure (see below) instead of comparing fits to individual subject data, because the high noise in these individual data leads to low confidence about the corresponding individual parameters.

Data inclusion criteria.

We performed outlier rejection on the learning curves of each experiment. Specifically, for each trial, we excluded adaptation levels that were more than 3 IQRs away from the subject median. This resulted in the inclusion of 99.4% of trials. Moreover, 1 participant in Experiment 3 was excluded from analysis due to inability to follow the experimenter's instructions.

Estimation of visuomotor rotation adaptation.

To assess the amount of adaptation to the trained VMR, we measured the direction of hand motion on each trial. In movements with visual feedback, this was defined as the direction of the vector between the hand position at movement onset (based on a 6.4 cm/s velocity threshold) and the hand position 150 ms later. We used 150 ms to measure feedforward adaptation, as feedback corrections should be minimal at this point. In movements with no visual feedback used to estimate temporally-persistent adaptation and 24-h retention in Experiment 4, this was defined as the direction of the vector between the hand position at movement onset and the movement endpoint.

To examine learning-related changes in performance, we subtracted out the small bias present in the baseline (0.13 ± 0.11°) from all the movement-direction data.

Measurement of temporally-persistent and temporally-volatile adaptation.

In Experiments 1, 2, 3, and S1, we measured temporally-persistent adaptation using 1-min delays interspersed with training. Because the temporally-volatile component of motor adaptation decays with a time constant of 15 to 25 s [56–58], the 1-min delays we impose here amount to 2.5 to 4τ, and thus lead to approximately 95% decay in temporally-volatile adaptation, effectively isolating the temporally-persistent component of adaptation. In contrast, the trial-to-trial decay in temporally-volatile adaptation for non-delay trials would be much lower, as the experiments were fast-paced with a median inter-trial time interval of 2.5 to 2.7 s, amounting to 0.1 to 0.2τ, thus leading to only 10% to 15% decay.

Thus, adaptation on the trial immediately following such a delay was operationally defined as temporally-persistent adaptation (**Fig 1D**). The corresponding overall adaptation as operationally defined as the average adaptation 2 trials before and 2 trials after the post-delay trial (with the exception of Experiment 3 which had additional trials to further dissociate adaptation into implicit and explicit components; see Experiment protocol section above). Temporally-volatile adaptation was taken as the difference between overall and temporally-persistent adaptation (**Fig 1D**).

These timed 1-min delays occurred every 30 trials during the VMR training blocks (on trials 10, 40, and 70 after the onset of each 80-trial training episode) and in 40-trial intervals during the long washout period, as shown in **Fig 1B and 1C**. During these delays, subjects held the handle still on the starting position. In addition to these timed 1-min delays, the Experiments contained rest breaks that allowed subjects to put the handle aside and were not strictly timed. These breaks occurred only during baseline or washout periods as shown in **Fig 1C**. We used the amount of adaptation after these breaks as a measure of temporally-persistent adaptation, but only when these breaks amounted to inter-trial intervals greater than 40 s (65.8% of these breaks for Experiments 1 to 3).

In Experiment 4, temporally-persistent adaptation was assessed during the no-feedback testing blocks that followed rest breaks (average break duration: 125 ± 8 s, minimum 58 s). Given a time constant for the decay of the temporally-volatile component of 15 to 25 s [56–58], this break would allow >99% decay in temporally-volatile adaptation and thus isolate temporally-persistent adaptation. Temporally-persistent adaptation was measured as the

average of 6 reaches (3 reaches in each of the 2 no-feedback testing blocks) that were towards the training target. We designed Experiment 4 with 2 test blocks because we thought that averaging the data from both blocks might reduce the effects of measurement noise, as we expected that both temporally-persistent measurements would predict 24-h retention, but that the average might make a cleaner prediction. We estimated volatile adaptation as the difference between persistent adaptation and overall adaptation. The latter was assessed as the average adaptation during the last 20 trials of the training and retraining blocks. Finally, we calculated 24-h retention based on the no-feedback data from the testing block on day 2 (average of 6 reaches to the previously trained target, split over 2 consecutive blocks, **Fig 5A**).

Estimation of washout time constants.

The washout of overall adaptation proceeded in 2 timescales: a very rapid initial washout phase during the first 2 to 3 washout trials, during which adaptation levels went from about 27° to about 11°, and then a slower washout phase that is illustrated in **Fig 1C**. To compare the time constants for washout for both temporally-persistent and overall adaptation (**Fig 2A**), we focused our overall washout analysis on the period beginning at trial 3 of washout in order to focus on the slower washout phase for comparing overall and temporally-persistent washout, because no temporally-persistent measurements were available during the very fast initial phase.

To estimate the values and confidence intervals associated with the time constants for washout, τ, we utilized a bootstrapping procedure [95]. Specifically, for each one of 10,000 bootstrap iterations, we randomly sampled, with replacement, $N = 20$ subjects from each group, and fit their average data with a single-exponential fit (Eq 1):

$$\boxed{?}$$

<div align="right">(1)</div>

When analyzing the overall washout curves, we discarded not only trials after each 1-min or rest break that removed the temporally-volatile component of adaptation in order to measure temporally-persistent learning, but also the 3 trials immediately thereafter, during which temporally-volatile adaptation might not be fully reequilibrated.

Normalization of adaptation data.

To systematically quantify savings, and specifically take into account the systematically different baselines between post-long washout versus post-short washout relearning, as well as the different baselines between temporally-persistent, temporally-volatile, and overall adaptation, we

subtracted baseline adaptation, , and normalized each learning curve *x* by the distance between baseline and the ideal adaptation level of 30 degrees (Eq 2). The baseline level for overall adaptation was defined as the average of the last 5 trials before training onset, whereas the baseline level for persistent adaptation was defined as the average of the last 3 persistent-adaptation trials before training onset (in the case of baselines for initial training and training after an 800-trial washout) or as the last single persistent-adaptation trial, trial, 10 trials before the onset of training (in the case of baselines for training after a 40-trial washout, since the 40-trial washout contained only a single persistent-adaptation measurement trial).

(2)

Estimation of savings.

Finally, savings for each type of adaptation were taken to be the % difference in adaptation between initial training and retraining for the same training trial (Eq 3).

(3)

Throughout the study, we focused on savings around 1-min delay trials (especially trial 10 after training onset which captured early adaptation, but also trials 40 and 70), as these were the trials for which all 3 types of adaptation could be assessed. For the analysis of the across-individual relationships between savings and persistent/volatile adaptation (**Fig 6C and 6D**), however, because the measurement of temporally-volatile adaptation was based on the same measurements as overall adaptation (volatile = overall [adaptation 2 trials before and after the 1-min delay trial]–persistent [adaptation on the 1-min delay trial]), we instead calculated overall savings based on trials 2 to 6 (relative to rotation onset) in order to ensure that any observed relationships were not due to measurements shared between the dependent (savings) and independent (temporally-volatile adaptation) variables. This range was selected as it was both relatively far from the measurements used to calculate temporally-volatile adaptation, but also better captured the rapid rise of overall adaptation providing more power to assess inter-individual differences in savings.

Comparisons of inter-individual differences.

To examine contributions of temporally-persistent or temporally-volatile adaptation on savings and long-term memory (**Fig 6A–6D**), we used linear regression with slopes restricted to positive values to model positive contributions of these components of adaptation and either savings or long-

term retention. Specifically, for studying long-term memory, we compared temporally-persistent and temporally-volatile adaptation on day 1 in Experiment 4 against 24-hour retention on day 2, whereas, for studying savings, we compared temporally-persistent and temporally-volatile adaptation from trial 10 in the retraining blocks in Experiments 1 and 2 against overall savings calculated as in the preceding paragraph.

Supporting information

Anti-savings in temporally-persistent adaptation cannot be explained by prolonged reaching under baseline conditions.

-800
-220
0
0
10
20
30
10
40
70
80
Adaptation (degrees)
Initial learning after 220 baseline trials
Relearning after 800 washout trials
Intial learning after 800 baseline trials (new data)
From Experiments 1 and 2 combined
Trial number relative to learning onset
-10
01
02
03
04
05
06
07
08
0
0
10
20
30
Adaptation (degrees)
Overall
Volatile
Persistent
0
20
40
60
80
100
Adaptation (%)
Trial 10 adaptation
Overall
Volatile
Persistent
0
20
40
60
80
100
Trial 40 adaptation
Overall
Volatile
Persistent
0
20
40
60
80
100
Trial 70 adaptation
Trial number relative to learning onset
*
NS

a

c

b

Comparison of new data to initial learning and relearning data from Experiments 1 / 2
Zoomed-in version of above focusing on the training period
Direct comparison of persistent, volatile, and overall adaptation levels
**
NS
*

S1 Fig.

Anti-savings in temporally-persistent adaptation cannot be explained by prolonged reaching under baseline conditions.

Here we investigated whether the anti-savings found in Experiments 1 / 2 (most pronounced following an 800-trial washout) might be due to the prolonged 800-trial washout period strengthening the baseline, unadapted state to the point that it resists the formation of a temporally-persistent memory of adaptation during relearning. Previous literature suggests that this kind of repetition effect – a form of use-dependent learning (Classen et al., 1998; Diedrichsen et al., 2010) – tends to level off after only 50-150 trials (Verstynen & Sabes, 2011); thus, it should equally affect initial learning (which follows 220 baseline trials) and relearning after 800 washout trials, suggesting no net effect in the anti-savings we observe. However, this use-dependent learning effect has not been studied within the specific context of our task. Thus, in Experiment S1, we examined 12 new participants who adapted to a 30° visuomotor rotation following an 800-trial baseline period, to match the long washout period in Experiments 1 and 2. We found that the prolonged baseline in Experiment S1 (light gray) did not reduce the temporally-persistent component during adaptation compared to the shorter, 220-trial baseline in Experiments 1 and 2 (dark gray); instead, relearning after a 800-trial washout in Experiments 1 and 2 led to significant reductions in temporally-persistent adaptation, as we discuss in the main text. Together, these findings show that anti-savings in temporally-persistent adaptation were not due to the use-dependent learning during the long 800-trial washout period.

(a)

 Comparison of average adaptation curves for (i) initial learning after 800 baseline trials from Experiment S1 (light gray), (ii) initial learning after 220 baseline trials from Experiments 1 / 2 (dark gray), and (iii) relearning after 800 washout trials from Experiments 1 / 2 (blue).

(b)

Close-up of the adaptation phase, with temporally-persistent measurements indicated by the empty circles as in Figure 3A. Note the similarity in the adaptation curves for the both initial learning cases (both after 220 and 800 trials of baseline) in contrast to the relearning curve. Errorbars indicate SEM; red lines indicate 60-second delays used to isolate temporally-persistent adaptation.

(c)

Comparison of the levels of overall, temporally-persistent, and temporally-volatile adaptation for these three cases, at trials 10, 40, and 70 after the onset of the visuomotor rotation perturbation. Temporally-persistent adaptation displays no signs of reduction after the 800-trial baseline (new data) relative to the 220-trial one (Exp. 1/2 data); however, it is significantly higher than temporal-

ly-persistent adaptation during relearning after the 800-trial washout in Exp. 1/2. * $p<0.05$; ** $p<0.01$. Underlying data supporting this figure can be found in files Exp_1_2_data.mat and Exp_S1_data.mat.

S1 Fig. **Anti-savings in temporally-persistent adaptation cannot be explained by prolonged reaching under baseline conditions.**

Here, we investigated whether the anti-savings found in Experiments 1/2 (most pronounced following an 800-trial washout) might be due to the prolonged 800-trial washout period strengthening the baseline, unadapted state to the point that it resists the formation of a temporally-persistent memory of adaptation during relearning. Previous literature suggests that this kind of repetition effect—a form of use-dependent learning [69,70]—tends to level off after only 50–150 trials [71]; thus, it should equally affect initial learning (which follows 220 baseline trials) and relearning after 800 washout trials, suggesting no net effect in the anti-savings we observe. However, this use-dependent learning effect has not been studied within the specific context of our task. Thus, in Experiment S1, we examined 12 new participants who adapted to a 30° visuomotor rotation following an 800-trial baseline period, to match the long washout period in Experiments 1 and 2. We found that the prolonged baseline in Experiment S1 (light gray) did not reduce the temporally-persistent component during adaptation compared to the shorter, 220-trial baseline in Experiments 1 and 2 (dark gray); instead, relearning after an 800-trial washout in Experiments 1 and 2 led to significant reductions in temporally-persistent adaptation, as we discuss in the main text. Together,

these findings show that anti-savings in temporally-persistent adaptation were not due to the use-dependent learning during the long 800-trial washout period. **(a)** Comparison of average adaptation curves for (i) initial learning after 800 baseline trials from Experiment S1 (light gray), (ii) initial learning after 220 baseline trials from Experiments 1/2 (dark gray), and (iii) relearning after 800 washout trials from Experiments 1/2 (blue). **(b)** Close-up of the adaptation phase, with temporally-persistent measurements indicated by the empty circles as in Fig 3A. Note the similarity between the adaptation curves for the initial learning cases (after 220 and 800 trials of baseline) in contrast to the relearning curve. Error bars indicate SEM; red lines indicate 60-s delays used to isolate temporally-persistent adaptation. **(c)** Comparison of the levels of overall, temporally-persistent, and temporally-volatile adaptation for these 3 cases, at trials 10, 40, and 70 after the onset of the visuomotor rotation perturbation. Temporally-persistent adaptation displays no signs of reduction after the 800-trial baseline (new data) relative to the 220-trial one (Exp. 1/2 data); however, it is significantly higher than temporally-persistent adaptation during relearning after the 800-trial washout in Exp. 1/2. $*p < 0.05$; $**p < 0.01$. Underlying data supporting this figure can be found in files Exp_1_2_data.mat and Exp_S1_data.mat.

https://doi.org/10.1371/journal.pbio.3001799.s001